5633076

```
                                                                    sau96I
                                                                    haeIII
                                                                    asuI
                                                                    sau96I
                                                                    nlaIV
                                                                    hgiJII
                                                                    ecoO109I
                                                                    bsp1286
                                                                    banII
                                                                    asuI
                                                                    nlaIV
                                mnlI                           aval apaI mnlI
                      mboII                              mnlI ecoO109I        tth111I
    1    GGA CTT GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG
  -18    Gly Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu


                      haeIII
                      eaeI                                        hInPI
                      cfrI                                        hhaI
   46 TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC
   -3 Cys Leu Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser


                      haeIII
                      mnlI
                 aval haeI
   94 CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AAA
   14 Gln Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys


                 mnlI            fnu4HI
             sau96I             bbvI                    pleI
             haeIII             aluI                    hInfI            bsrI
             asuI               pvuII              bsmaI        fokI
  142 GTG CTG GGC CCT CCT GTC AGC TGC ATA AAG AGA GAC TCC CCC ATC CAG
   30 Val Arg Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln


                 haeIII
                 haeI
              scrFI                      haeIII
              ecoRII                     sau96I
              bstNI                      asuI     sfaNI
  190 TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT
   46 Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp


                                    sau96I
                                    nlaIV
                                 scrFI
                                 ecoRII
                                 bstNI
                                 haeIII
                              stuI     haeIII
                   mnlI  haeI     asuI
  238 GGT GGT TTC ATA TAC GAG GCA GGC CTG GCC CCC TAC AAA CTG CGA CCT
   62 Gly Gly Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro
```

FIG.—1A

```
                                    sau96I
                                    avaII
                                    asuI
          fnu4HI         accI      nlaIV
286 GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT
 78 Val Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr


                                        fnu4HI
                            mboII        bbvI          aluI
                    hphI           fnu4HI aluI     pvuII
334 TAT GCC GTG GCT GTG GTG AAG AAG GGC GGC AGC TTT CAG CTG AAC GAA
 94 Tyr Ala Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu


                                    haeIII        sau96I
                                    stuI          avaII
                        bglI        haeI          asuI          fokI
382 CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA
110 Leu Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly


                                                            sau96I
                                                            avaII
                                                            asuI
                                                            nlaIV
430 TGG AAT GTC CCT ACA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT
126 Trp Asn Val Pro Thr Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly


          hgiJII           aluI
          bsp1286          fnu4HI
          banII            bbvI                    ddeI       aluI
          ddeI        mnlI pvuII       mboII           pvuII
478 CCA CCT GAG CCC ATT GAG GCA GCT GTG CAG TTC TTC TCA GCC AGC TGT
142 Pro Pro Glu Pro Ile Glu Ala Ala Val Gln Phe Phe Ser Ala Ser Cys


     mspI
     hpaII
    scrFI
    nciI
   cauII
526 GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT
158 Val Pro Gln Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys


                                        nlaIV
                                       scrFI
                                       ecoRII
                               mnlI bstNI        rsaI
574 GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC TCC TCC CAG GAA CCG TAC
174 Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr
```

FIG.-1B

```
                                 nlaIV
                                 hgiCI
         aluI                    banI                    ddeI  bsmaI              bsmaI
622 TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA GAC
190 Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp


                                                         sau96I
                                                         avaII
                                                         asuI
                                                         ppuMI
                                 hgiAI                   ecoO109I
                                 bsp1286           mnlI            mnlI
670 GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT GAG GAC CTG TCA GAC GAG
206 Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu


718 GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG
222 Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys


                                                         scrFI
                                                         nciI
                                                         mspI
                                                         hpaII
                                                         cauII
                                                    xmaI sau96I
                                                    smaI nlaIV
                                                    scrFI
                                                    nciI avaII
                                                    cauII
                                                    avaI asuI
                                                sau96I ppuMI
                                                haeIII nlaIV
      bsrI                                       asuI    ecoO109I         nlaIII
766 CCA GTG GAC AAG TTC AAA GAC TGC CAT CTG GCC CGG GTC CCT TCT CAT
238 Pro Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His


                                                    sfaNI
                                                    fokI                mboII
      bglI              draIII                      mnlI           hinfI
814 GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG AAT
254 Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn


         scrFI
         ecoRII
         bstNI                                                hphI
862 CTT CTC CGC CAG GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG AAA
270 Leu Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys
```

FIG.-1C

```
                                                              sau3AI
                                                              mboI
                                                              dpnI
                                                              xhoII
           aluI                                               bstYI
           bstXI                    nlaIV                      bglII
       910 TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG TTC AAG
       286 Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys


                                       nlaIV
                                       hgiCI
           pleI                    mnlI bsp1286    mnlI
           hinfl                      taqI banI      avaI              hinfl
       958 GAC TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG
       302 Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly


                          mspI
              styI        hpaII
           rsaI     nlaIV                           fokI              mnlI
      1006 CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC CAG AAC TTG AGG AAA
       318 Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys


                              mspI
                              hpaII        thaI
                              scrFI        fnuDII
                              nciI         bstUI
                   mnlI       fnu4HI       hinPI
            mnlI              bbvI cauII   hhaI
      1054 AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT GCG
       344 Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala


                              hinPI
                              mstI
                              fspI
                              fnu4HI
                              aluI hhaI         bstXI
                        alwNI bbvI              bsrI
      1102 GTG GGC GAG CAG GAG CTG CGC AAG TGT AAC CAG TGG AGT GGC TTG AGC
       350 Val Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser


                fnu4HI                        mnlI
                bbvI       bspMI    mnlI haeIII          mnlI     sfaNI
      1150 GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC
       366 Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile


                scrFI
                ecoRII                        bstXI
                bstNI                aluI sfaNI nlaIII     fokI   mnlI
      1198 GCC CTG GTG CTG AAA GGA GAA GCT GAT GCC ATG AGT TTG GAT GGA GGA
       382 Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly
```

FIG.-1D

```
                         nlaIII                    nlaIV        scrFI
                          sphI                     hgiCI        ecoRII
              rsaI       nspClx                    banI         bstNI
         1246 TAT GTG TAC ACT GCA TGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG
          398 Tyr Val Tyr Thr Ala Cys Lys Cys Gly Leu Val Pro Val Leu Ala Glu


                                               sau3AI
                                               mboI
                                               dpnI
                                               alwI
         1294 AAC TAC AAA TCC CAA CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG GAT
          414 Asn Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp


                                                           sau3AI
                                                           mboI
              ecoNI            ecoRV                        dpnI
         1342 AGA CCT GTG GAA GGA TAT CTT GCT GTG GCG GTG GTT AGG AGA TCA GAC
          430 Arg Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp


              scrFI
              ecoRII
              bstNI
         1390 ACT AGC CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC ACC
          446 Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr


                                                   haeIII
                                                nlaIII
                                                styI sau96I          mboII
                              pstI              ncoI asuI            earI
         1438 GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC TTC
          462 Ala Val Asp Arg Thr Ala Gly TrP Asn Ile Pro Met Gly Leu Leu Phe


              nlaIV
              hgiJII
              bsp1286
              banII                              sspI            aluI bsp1286
         1486 AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT
          478 Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys


                            sau3AI
                            mboI
                            DpnI
              scrFI         xhoII
              ecoRII        bstYI            hgiAI
              bstNI    aval bglII            bsp1286
         1534 GCC CCT GGG TCT GAC CCG AGA TCT AAT CTC TGT GCT CTG TGT ATT GGC
          484 Ala Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly
```

FIG.-1E

```
                          hphI                                    bsp1286
      1582 GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAT GAG AGA TAC
       510 Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr


                                        nlaIV
                                        hgiCI
                                     banI scrFI
                                  mspI      ecoRII
                        bsrI     hpaII bstNI ddeI bsmI          bsmaI
      1630 TAC GGC TAC ACT GGG GCT TTC CGG TGC CTG GCT GAG AAT GCT GGA GAC
       526 Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp


      1678 GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT
       542 Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn


                                                                 fnu4HI
                                                                 bbvI
                                                                 hinPI
                     mnII nlaIII ddeI              aluI          hhaI
      1726 AAC AAT GAG GCA TGG GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG
       558 Asn Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu


                    taqI                                         fnu4HI
                  mnII                              mnII         bbvI
                 bglI                              ddeI         aluI
      1774 CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC TGC
       574 Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys


                      sau96I
                      nlaIV
                    nlaIII
                  styI haeIII
                  ncoI asuI      hinfI nlaIII      bsmaI  fokI
      1822 CAT CTT GCC ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG
       590 His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys


                                      fnu4HI
                  ecoNI        alwNI bbvI
      1870 GTG GAA CGC CTG AAA CAG GTG CTG CTC CAC CAA CAG GCT AAA TTT GGG
       606 Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly


                  sau3AI
                  mboI            mspI
                  dpnI            hpaII
                  xhoII           scrFI
                  bstYI           nciI
                  alwI            cauII                          bsrI
      1919 AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA
       622 Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu
```

FIG.–1F

```
                                                            haeIII
                                                            haeI
                                                            eaeI              styI
                                            ddeI            cfrI    pleI ncoI
                                            draIII          balI    hinfI
1966 ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG GCC AGA CTC
 638 Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu


                                              sau96I
                                              avaII
                                              asuI
        nlaIII            ndeI        sspI    nlaIV
2014 CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA
 654 His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala


                                                    scrFI
                                                    ecoRII
                                hgIAI               bstNI
                                bsp1286   mnlI    mnlI
2062 GGC ATT ACT AAT CGT AAA AAG TGC TCA ACC TCC CCC CTC CTG GAA GCC
 670 Gly Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala


            ddeI
            mstII
            mnlI                          sau96I
            eco81I               mboII    haeIII
       ecoRI bsu36I         mboII     asuI aluI
2110 TGT GAA TTC CTC AGG AAG TAA  AACCGAAGAA GATGGCCCAG CTCCCCAAGA
 685 Cys Glu Phe Leu Arg Lys □C*


                                                    styI
                                                    haeIII
                                                    sau96I
                         mboII    scrFI             asuI
        ddeI             earI     ecoRII            nlaIV
        mnlI             aluI     bstNI       ecoO109I nlaIV
2161 AAGCCTCAGC CATTCACTGC CCCCAGCTCT TCTCCCCAGG TGTGTTGGGG CCTTGGCTCC


     ecoNI                   fokI                           ddeI
2221 CCTGCTGAAG GTGGGGATTG CCCATCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA

2281 AGAAGTAAAA TGAGAAATTT TGTTGATATT CAAAAAAAA

>LENGTH: 2319
```

FIG.-1G

```
  1  GACTCCTAGG GGCTTGCAGA CCTAGTGGGA GAGAAGAAC ATCGCAGCAG CCAGGCAGAA CCAGGACAGG TGAGGTGCAG GCTGGCTTTC CTCTCGCAGC
101  GCGGTGTGGA GTCCTGTCCT GCCTCAGGGC TTTTCGGAGC CTGGATCCTC AAGGAACAAG TAGACCTGGC CGCGGGAGGT GGGGAGGGGAA GGGGTGTCTA
201  TTGGGCAACA GGGGCGGCAAA AGCCCTGAATA AAGGGGCGGA GGGCAGGCGCA CCTTCGTTTG CCAAGTCGCC AGAC ATG AAA CTT
-19                                                                                                 M   K   L

304  GTC TTC GTC CTC CTG CTC TTC CTC GGG GCC CTC GGA CTG GCT GGC CTC GCT GGG GTT CAG TGG TGC GTA TCC
-16   V   F   V   L   L   L   F   L   G   A   L   G   L   A   G   L   A   G   V   Q   W   C   V   S

391  CAA CCC GAG GCC ACA TGC TTC CAA TGG CAA AGG AAT ATG CGT CCT GGC CCT GTC AGC TGC ATA AAG AGA GAC TCC
 14   Q   P   E   A   T   C   F   Q   W   Q   R   N   M   R   P   G   P   V   S   C   I   K   R   D   S

478  CCC ATC CAG TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GGT GCA GGC CTG GCC
 43   P   I   Q   C   I   Q   A   I   A   E   N   R   A   D   G   A   G   L   A

565  CCC TAC AAA CTG CGA CCT GTA GCG GCG GAA GTC TAC GGG ACT CAC TAT CGC CCA CGA ACC CAC ... AAG
 72   P   Y   K   L   R   P   V   A   A   E   V   Y   G   T   H   Y   R   P   R   T   H   ...  K

652  AAG GGC GGC AGC TTT CAG CTG AAC GAA CTG CAA GGT CTG CAA GGC CTT CGC CGC ... GTG GTG AAG
101   K   G   G   S   F   Q   L   N   E   L   Q   G   L   Q   G   L   R   R   ...  V   V   K

739  ACA GGG ACA CTT CGT CCA TTC AAT TGG ACG GCT GCC CCT GAG CCC ATT GAG ... AGC TGT
130   T   G   T   L   R   P   F   N   W   T   A   A   P   E   P   I   E   ...  S   C

826  GTT CCC GGT GCA GAT AAA GGA TTC GCA CAG GGG TGT CGC CTG CTG TGT GAA AAC ... TCC CAG
159   V   P   G   A   D   K   G   F   A   Q   G   C   R   L   L   C   E   N   ...  S   Q

913  GAA CCG TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AAG GCT AGA GAG AGC ... GTG TTT
188   E   P   Y   F   S   Y   S   G   A   F   K   C   L   K   A   R   E   S   ...  V   F

1000 GAG GAC CTG TCA GAC GAG GCT GAG AGG GAC GAT GTT CTC CTG CCA GAC AAC ACT CGG AAG TTC AAA GAC
217   E   D   L   S   D   E   A   E   R   D   D   V   L   L   P   D   N   T   R   K   F   K   D

1087 TGC CAT CTG GCC CGG GTC CCT CAT GCC GTT GCA AGT GGC GAG AGT GAC GAA AGG ATC TGG AAT ... CGC CAG
246   C   H   L   A   R   V   P   H   A   V   A   S   G   E   S   D   E   R   I   W   N   ...  R   Q

1174 GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG CCG AGG ATA GAT TCT GGG CAG AAA GAT CTG TTC AAG GAC
275   A   Q   E   K   F   G   K   D   K   S   P   P   R   I   D   S   G   Q   K   D   L   F   K   D

1261 TCT GCC ATT GGG TTT TCG AGG GTG CCC GTC TCC CTT GGC TAC CTT GGC CTC TTC ACT GCC ATC CAG AAC CAG
304   S   A   I   G   F   S   R   V   P   V   S   L   G   Y   L   G   L   F   T   A   I   Q   N   L

1348 AGG AAA AGT GAG GAG GAA GTG GCG CGG GTC GTG TGG TGT GCG CGC AAG CGC AAG TGT AAC CAG
333   R   K   S   E   E   E   V   A   R   V   V   W   C   A   R   K   R   K   C   N   Q
```

FIG.-2A

```
1435 TGG AGT GGC TTG AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC GCC CTG GTG CTG AAA GGA GAA GCT
 362  W   S   G   L   S   E   G   S   V   T   C   S   S   A   S   T   T   E   D   C   I   A   L   V   L   K   G   E   A

1522 GAT GCC ATG AGT AGT TTG GAT GGA TAT GTG CCT TTG GTG CTG GCA GAG AAC TAC AAA TCC CAA
 391  D   A   M   S   S   L   D   G   Y   V   P   L   V   L   A   E   N   Y   K   S   Q

1609 CAA AGC AGT GAC GAT CCT AAC CCT GAT AGA CCT GTG GTT GCG GTT GTT AGG TCA GAC ACT AGC
 420  Q   S   S   D   D   P   N   P   D   R   P   V   V   A   V   V   R   S   D   T   S

1696 CTT ACC TGG AAC TCT GTG AAA GGC AAG TCC TGC AAG AAG GGC CTG CAC ATG GGC CTG CTC
 449  L   T   W   N   S   V   K   G   K   S   C   K   K   G   L   H   M   G   L   L

1783 TTC AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TTT TCA GAA TAT CAA AGC AGC CCT AGA TCT AAT CTC TGT GCT
 478  F   N   Q   T   G   S   C   K   F   D   E   F   S   E   Y   Q   S   S   P   R   S   N   L   C   A

1870 CTG TGT ATT GGC GAC GAG CAG AAT CAT AAG TGC GTG CCC AAC AGC AGA GAG AGA TAC ACT GAT GGA AAT AAC GAG GCT TGC
 507  L   C   I   G   D   E   Q   N   H   K   C   V   P   N   S   R   E   R   Y   T   D   G   N   N   E   A   W

1957 CTG GCT GAG AAT GCT CTG GCA GAC TTT GCG GCA GAC TGC GTC CTG AAA TTT GTG AAA GAT GTC ACT GAT GGC AAG TGC GCC
 536  L   A   E   N   A   L   A   D   F   A   A   D   C   V   L   K   F   V   K   D   V   T   D   G   K   C   A

2044 AAG GAT TTG AAC CTG GCA GAC TTT GCG GCA GAC CTC CTC CTG GAC GGG AAG CGG AAG AAG CAG GAA AGC CAT CTT CAT
 565  K   D   L   N   L   A   D   F   A   A   D   L   L   L   D   G   K   R   K   K   Q   E   S   H   L   H

2131 ATG GCC AAT CAT GCC GTG GTG GTG TCT CGG ATG GAT GAC AAG CTC AAC CTT CTT CAG CAG TAT TTT GGA CCA CAG TAT GGG
 594  M   A   N   H   A   V   V   V   S   R   M   D   D   K   L   N   L   L   Q   Q   Y   F   G   P   Q   Y   G

2218 AGA AAT GGA TCT GAC TGC CCG AAA GAC AAG TTT TGC GAC GAG AAG ACC TCT AAT GAC AAC CTG AAA AAG TGC TGT
 623  R   N   G   S   D   C   P   K   D   K   F   C   D   E   K   T   S   N   D   N   L   K   K   C   C

2305 GCC AGA GCC CTC CAT GGC AAA ACA GAT TTG GGA GCC GCC ATT ACT AAT CTG AAA GAG TGT TCA ACC
 652  A   R   A   L   H   G   K   T   D   L   G   A   A   I   T   N   L   K   E   C   S   T

2392 TCC CCC CTC CTG GAA GCC TGT GAA TTC CTC AGG AAG GAC TAA AACCGAGGAA GATGGCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA AGAAGTAAAA
 681  S   P   L   L   E   A   C   E   F   L   R   K   D   *

2491 TCTCCCCAGG TGTGTTGGGG CCTTGGCTCC CCTGCTGAAG GTGGGGATTG CCCATCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA AGAAGTAAAA
2591 TGAGAAATTT TGTTGATATT CAAAAAAAA
```

FIG.-2B

RESTRICTION MAP OF 5' - FLANKING REGION OF αS1 CASEIN GENE
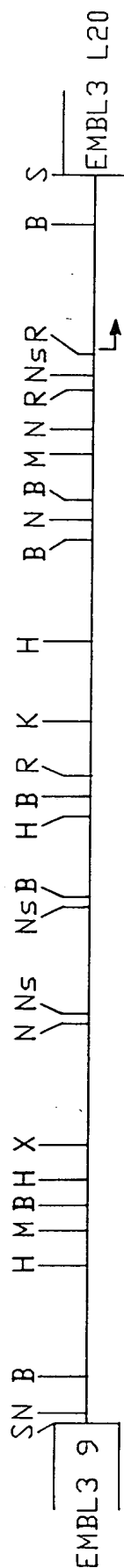
S – SalI

B – BalII

R – EcoRI

H – HindIII

M – SmaI

N – NcoI

Ns – NsiI

X – XhoI

K – KpnI

1Kb

EMBL3 9

EMBL3 L20

FIG.-3

RESTRICTION MAP OF 3' FLANKING
REGION OF α S1 CASEIN GENE
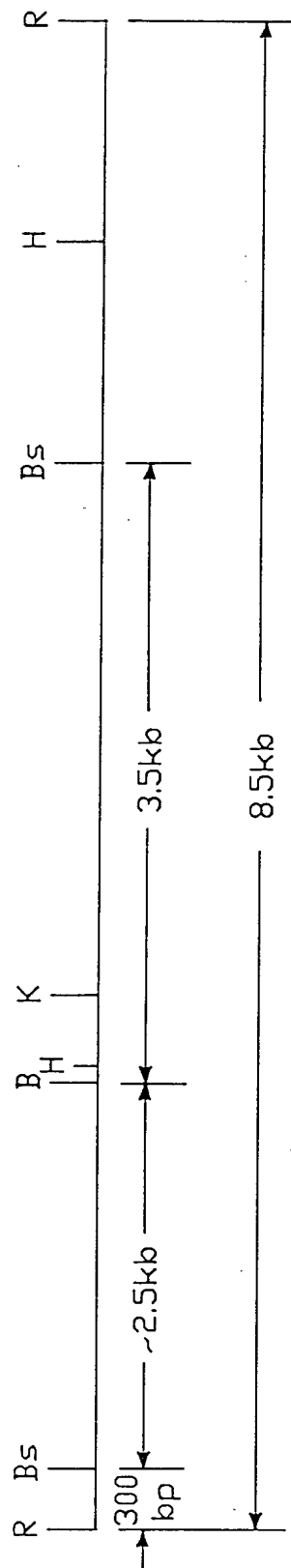
R — EcoRI

Bs — Bst EII

B — BamHI

H — HindIII

K — KpnI
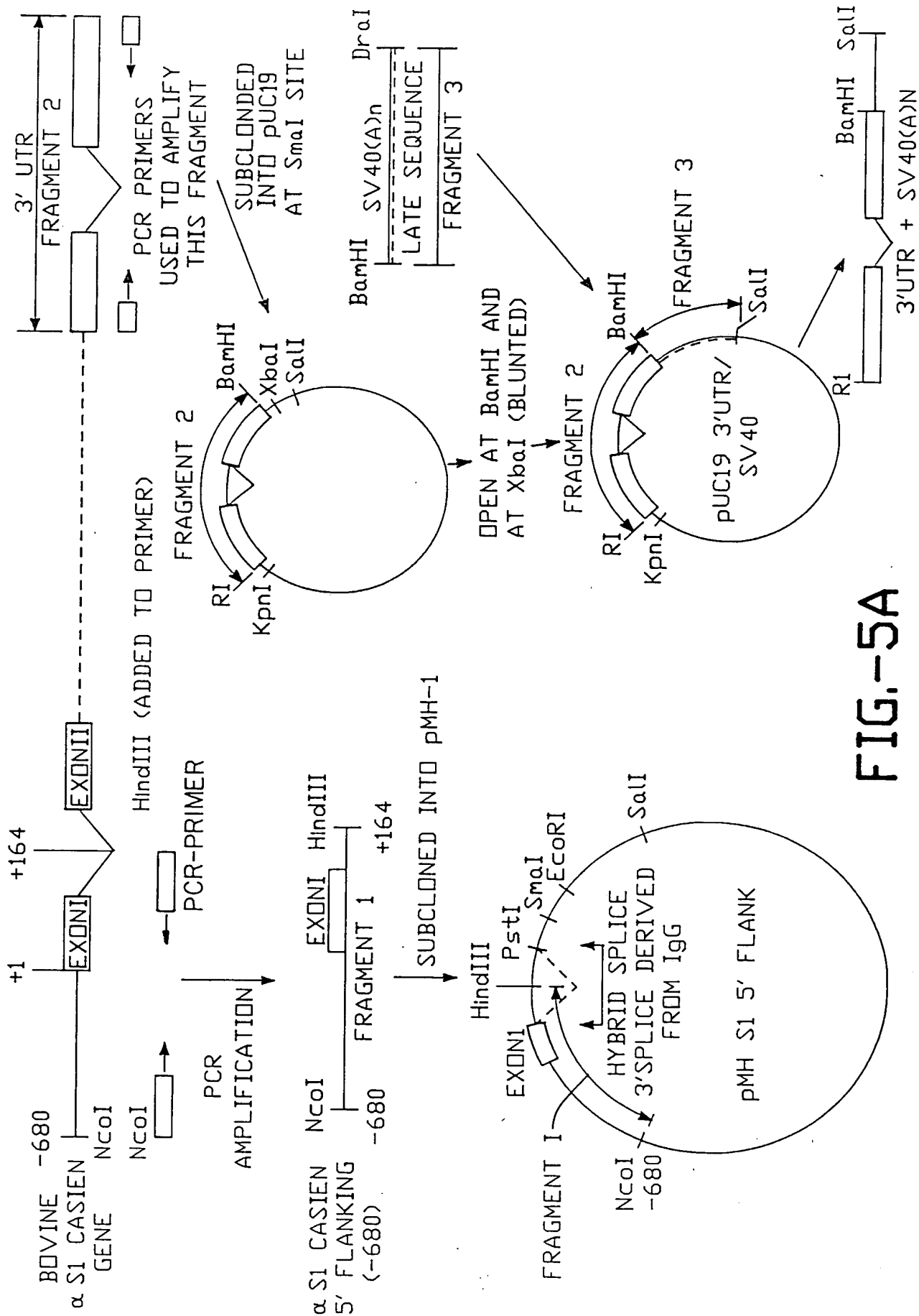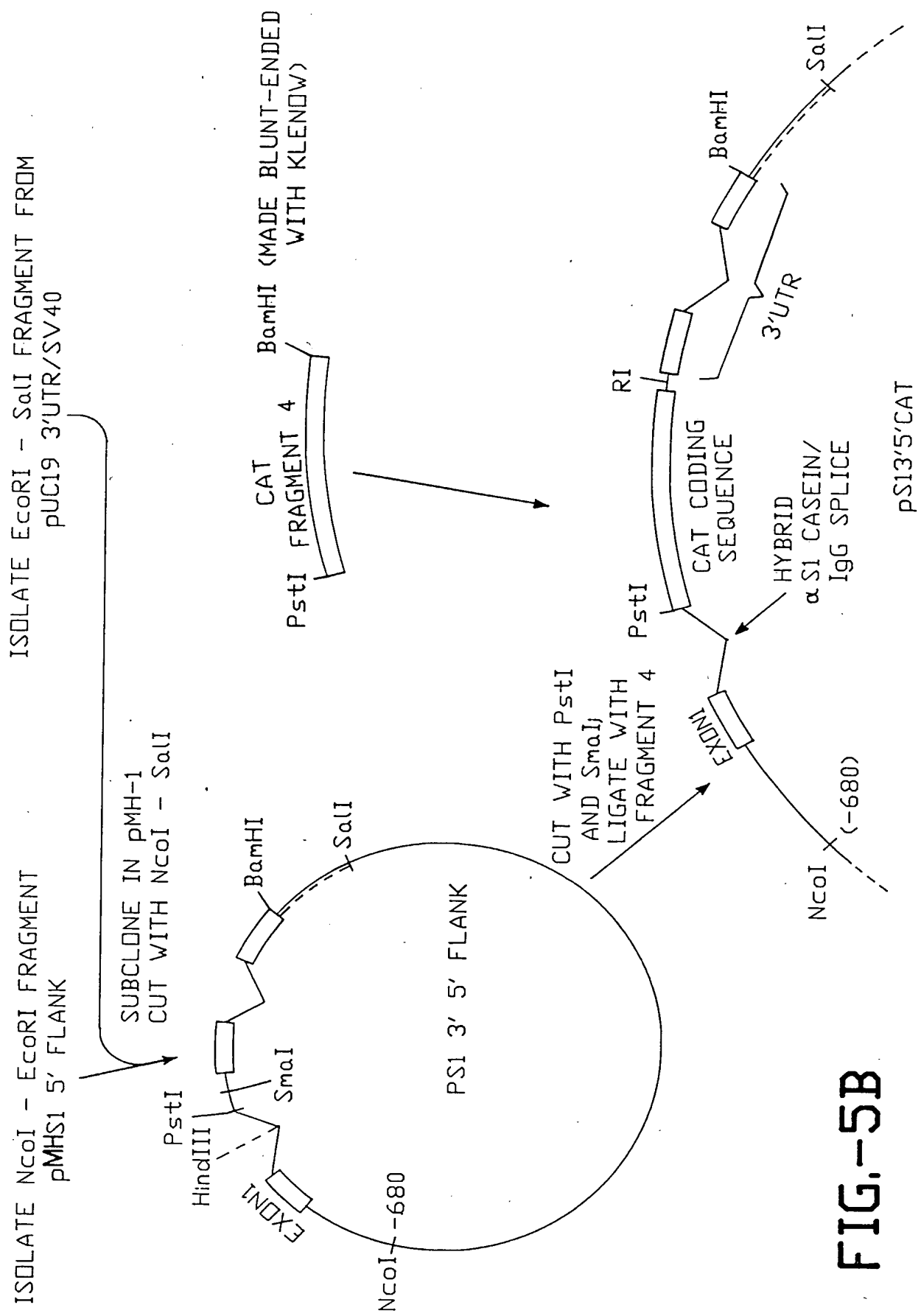
FIG.-4

FIG.-5A

ISOLATE NcoI – EcoRI FRAGMENT FROM pMHS1 5' FLANK

ISOLATE EcoRI – SalI FRAGMENT FROM pUC19 3'UTR/SV40

SUBCLONE IN pMH-1 CUT WITH NcoI – SalI

CAT FRAGMENT 4

BamHI (MADE BLUNT-ENDED WITH KLENOW)

PstI

CUT WITH PstI AND SmaI; LIGATE WITH FRAGMENT 4

PstI

HindIII

SmaI

EXON1

NcoI —680

PS1 3' 5' FLANK

BamHI

SalI

NcoI —680

EXON1

HYBRID αS1 CASEIN/ IgG SPLICE

(-680)

PstI

CAT CODING SEQUENCE

RI

3'UTR

BamHI

SalI

pS13'5'CAT

FIG.-5B

ISOLATE CAT + SV40(A)n LATE SEQUENCE

PstI   SV40   BLUNT

CAT CODING   (A)n LATE

CUT PstI/SmaI

PstI

CAT CODING

HYBRID
SPLICE

SV40(A)n

RI
SalI

EXON1

NcoI
-680

pS1 5'CAT

HindIII
PstI   SmaI
        RI

SalI

NcoI
-680

pMHS1 5'FLANK
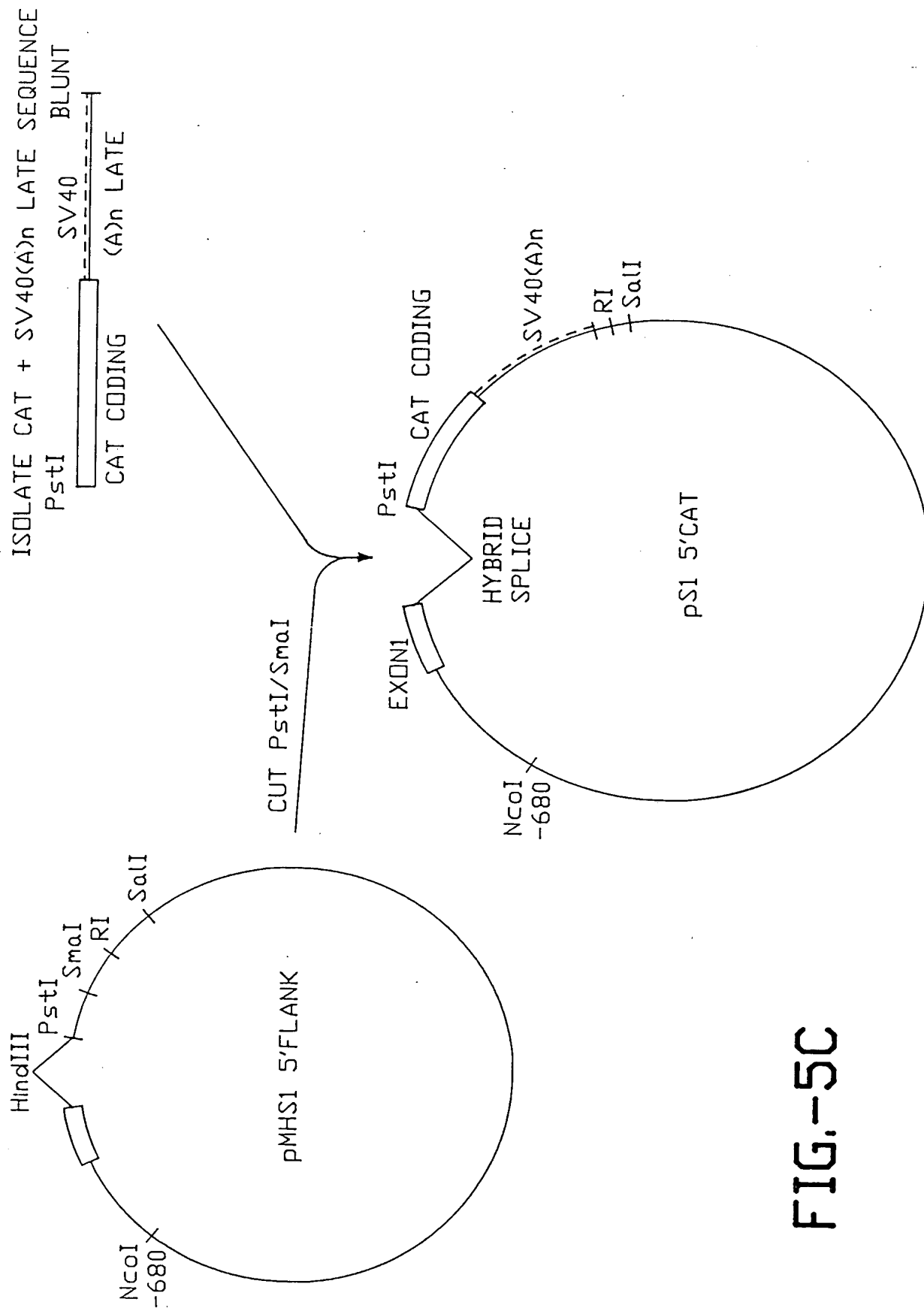
FIG.-5C

NAME pMH-1
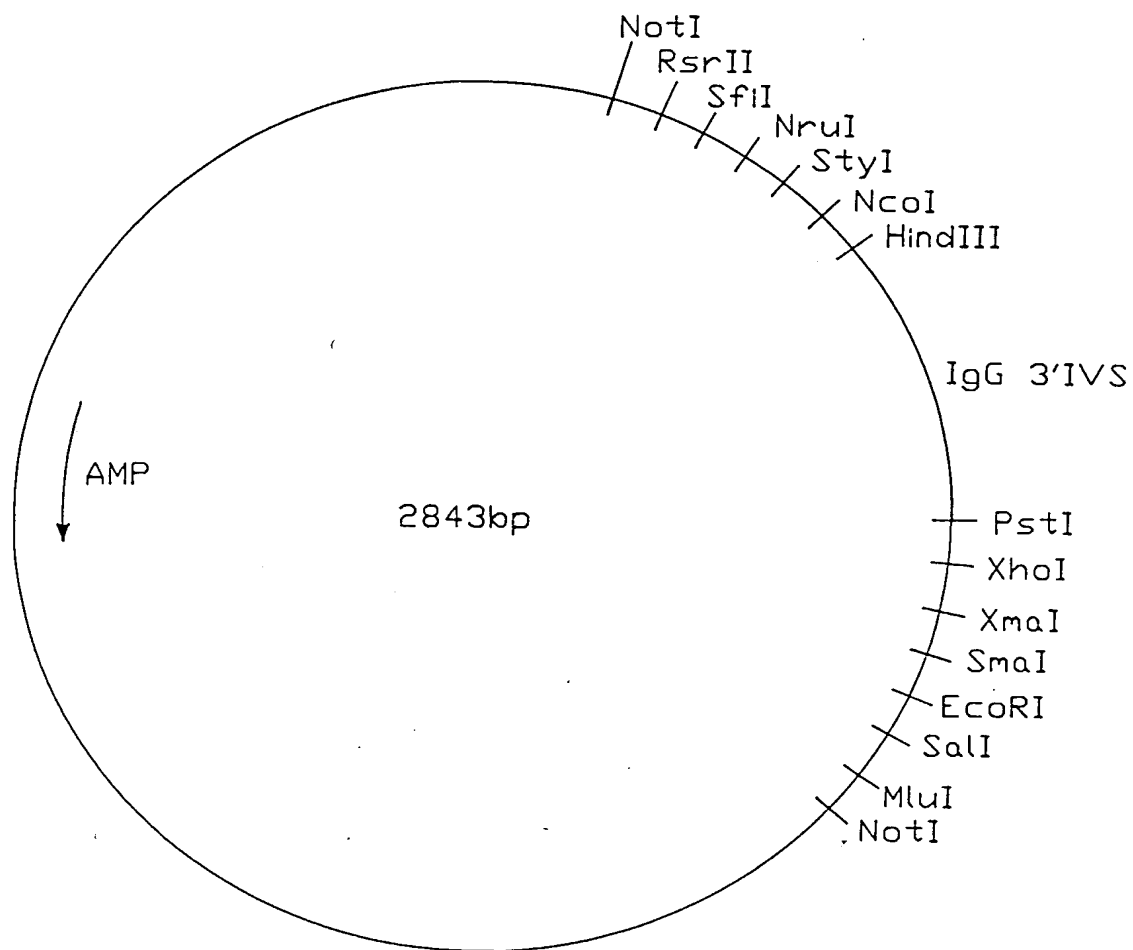
FIG.-6

- PLASMID pMH3'E10 CONTAINS ~ 8.5Kb EcoRI FRAGMENT OF THE 3'-END OF
  αS1 CASEIN. pMH-1 WAS CUT W/EcoRI + LIGATED TO THIS FRAGMENT

pMH3'E10

EcoRI
PstI
HindIII
NcoI
NotI

αS1 3'
FRAGMENT

EcoRI
SalI
NotI

EcoRI

pMH-1

$\xrightarrow[\text{DIGEST}]{\text{EcoRI}}$

8.5 Kb αS1 CASEIN
+ 3' EcoRI FRAGMENT

pMH3'UTRhLF2 LINKER
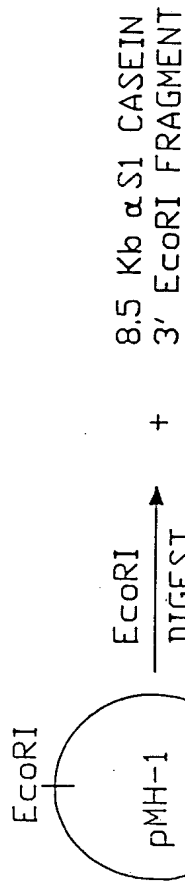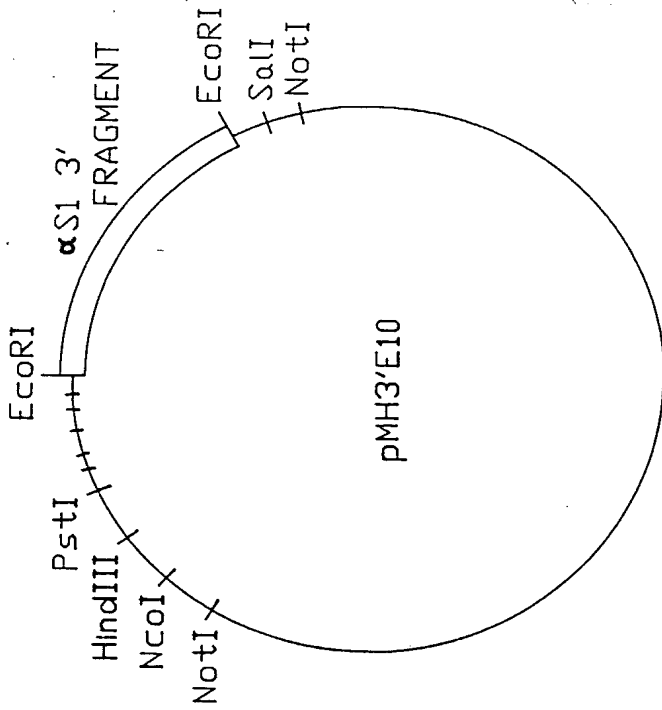
- PLASMID pMH3'UTRhLF2 LINKER

EcoRI

pMH3'UTR

KpnI

$\xrightarrow[\text{AND KpnI}]{\text{CUT W/EcoRI}}$
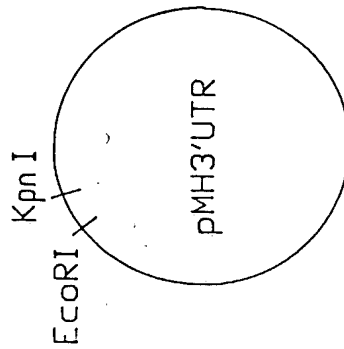
(FRAGMENT 7)
3' TERMINAL
hLF CODING
SEQUENCE
(~ 20bp)
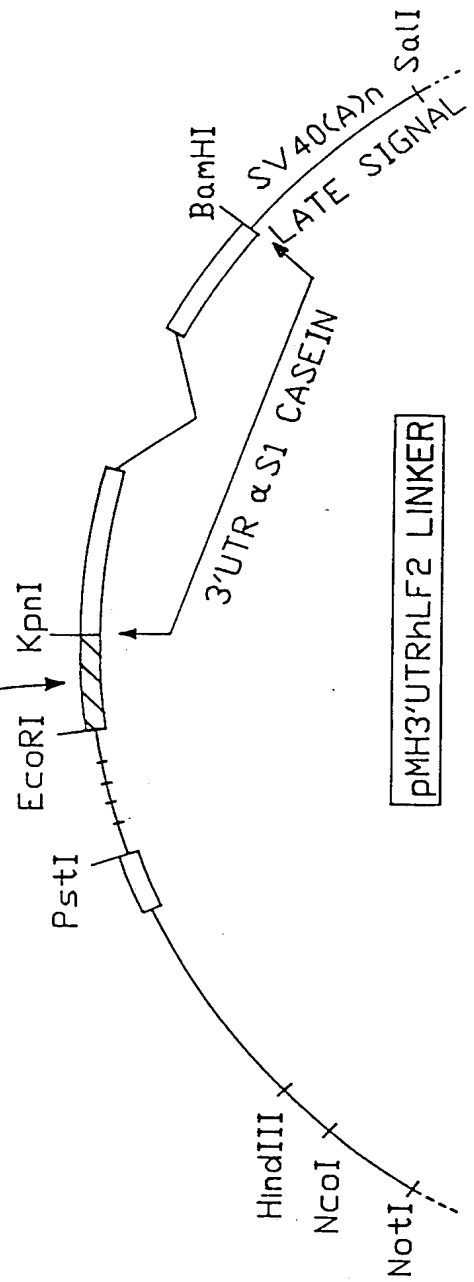
NotI
NcoI
HindIII

PstI

EcoRI
KpnI

3'UTR αS1 CASEIN

BamHI

SV40(A)n SalI
LATE SIGNAL

pMH3'UTRhLF2 LINKER

FIG.—7A

NotI HindIII
PstI
EcoRI KpnI
BamHI
SalI

pMH3'UTR
hLF2 LINKER

SV40(A)n

3'UTR αS1 CASEIN

NcoI

XhoI SmaI

CUT w/ XhoI + SmaI

ADD SYNTHETIC LINKER

αS1 SIGNAL SEQUENCE

XhoI

ClaI

NaeI

FRAGMENT 8

NaeI SITE (BLUNT ENDED)

αS1-CASEIN 3'UTR

SV40(A)n

BamHI

SalI
NotI

EcoRI KpnI
NaeI

αS1 SIGNAL

XhoI
ClaI
PstI

HindIII

NcoI
NotI
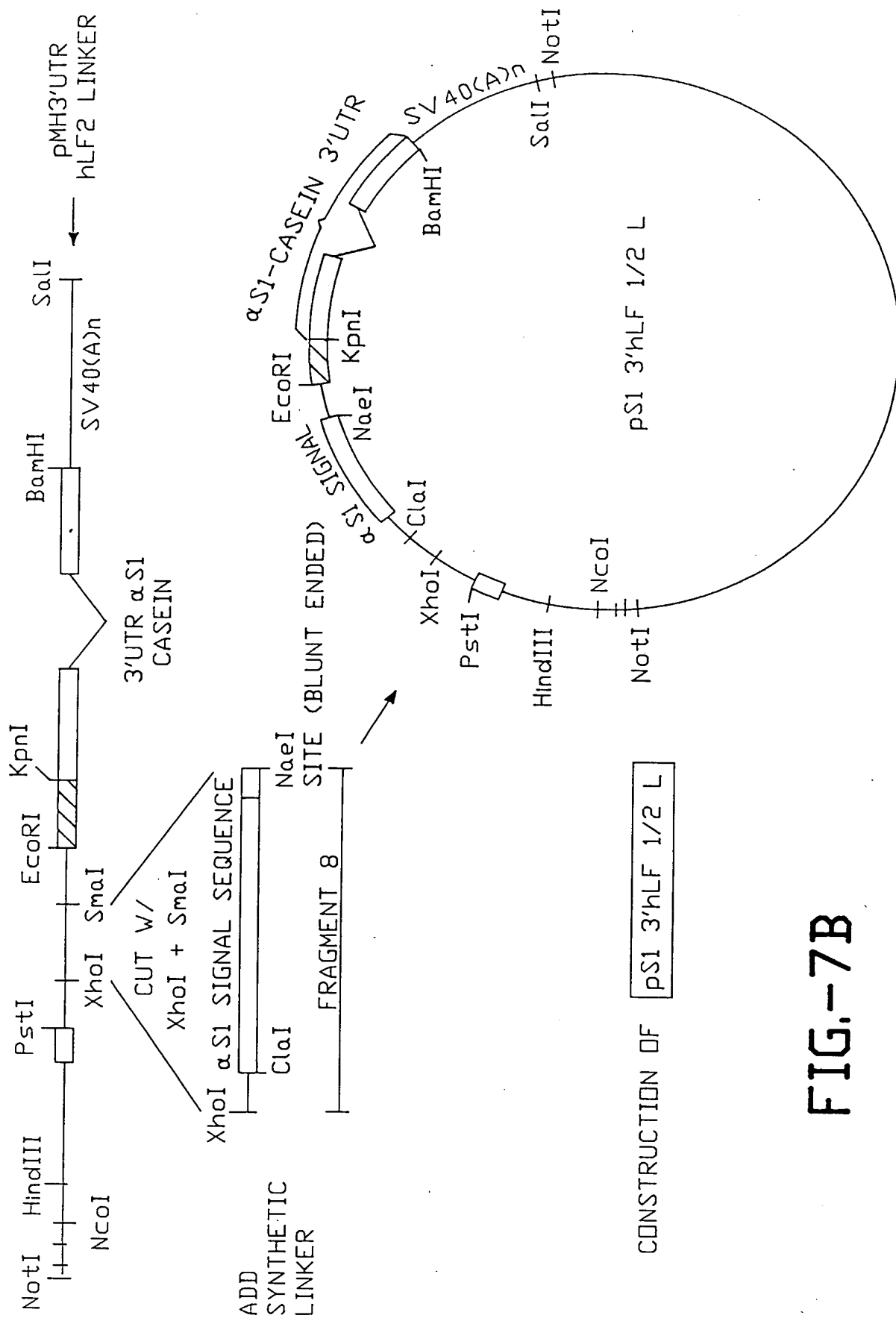
pS1 3'hLF 1/2 L

CONSTRUCTION OF pS1 3'hLF 1/2 L

FIG.–7B

CONSTRUCTION OF pS1 3'UTRhLF
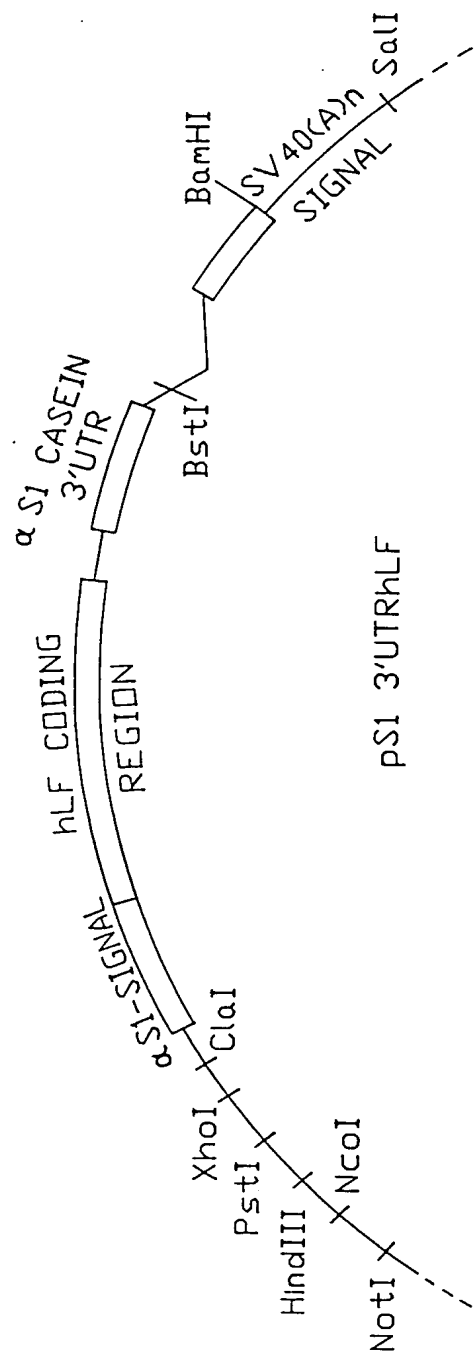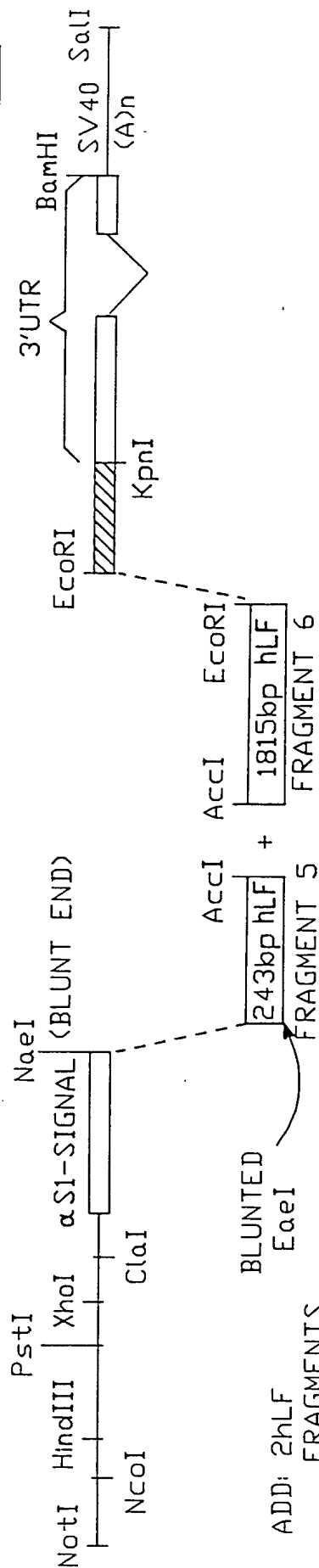
FIG.-7C

CONSTRUCTION OF phLF3' 10kb

STARTED W/pMH3'E10

→ ISOLATED THE BstEII TO SalI FRAGMENT

PURIFIED THE LARGE BstEII/SalI
FRAGMENT FROM pS1 3'UTRhLF AND
LIGATED THE BstEII/SalI FRAGMENT.

(~8.2kb)

LIGATED

phLF3' 10kb

α-S1-SIGNAL
hLF CODING
REGION
KpnI
BstEII
ClaI
XhoI
PstI
HindIII
NcoI
NotI
EcoRI
SalI

EcoRI
SalI
BstEII

FIG.-7D

CONSTRUTION OF pS1 3'5'hLF

STARTED W/pS1 3'UTRhLF. CUT W/HindIII AND SalI AND PURIFIED THIS FRAGMENT CONTAINING THE αS1-CASEIN SIGNAL SEQUENCE, hLF CODING REGION, αS1UTR AND SV40(A)n.

HindIII   PstI   hLF CODING   3'UTR   SV40(A)n   SalI

αS1-SIGNAL
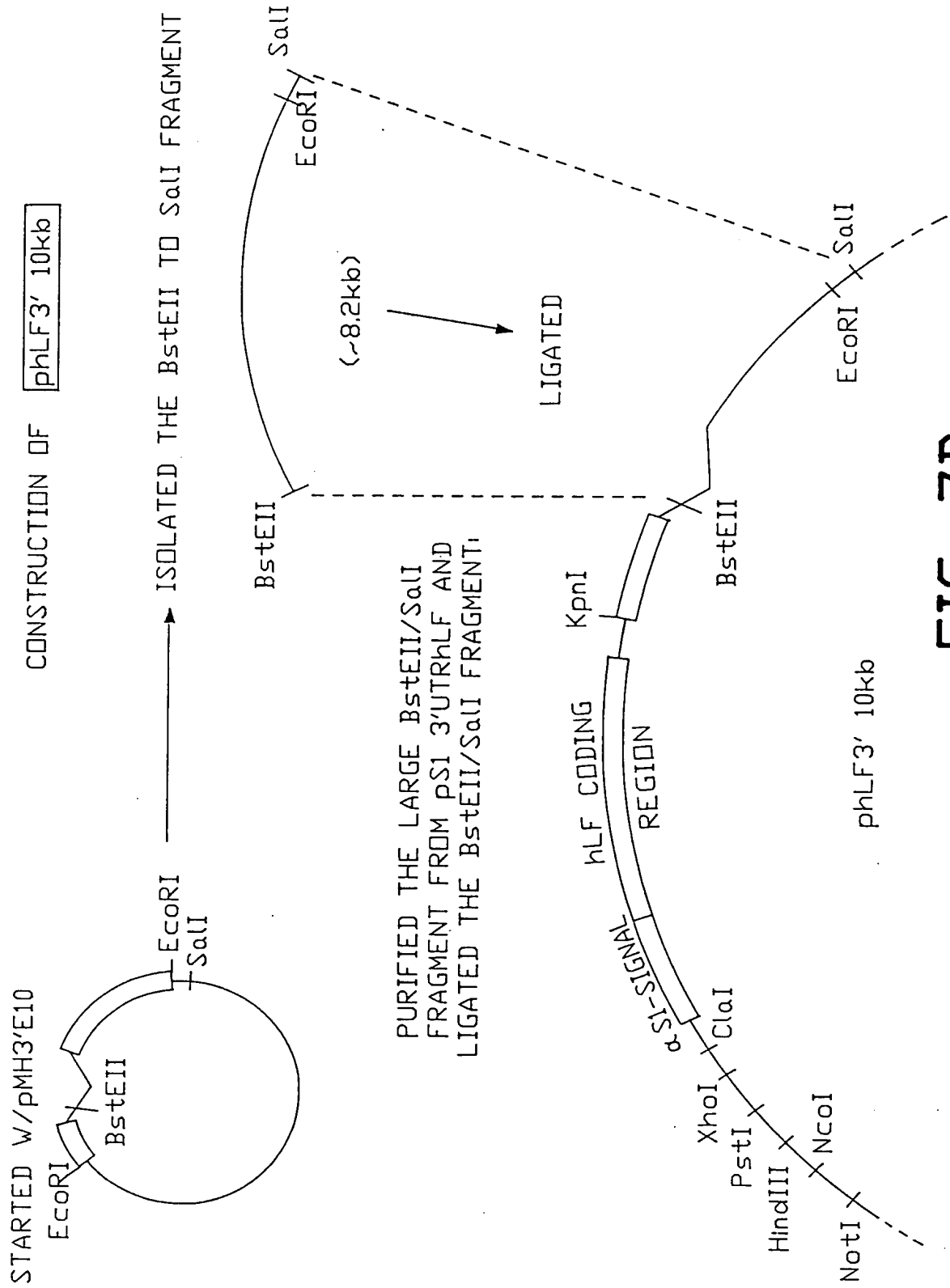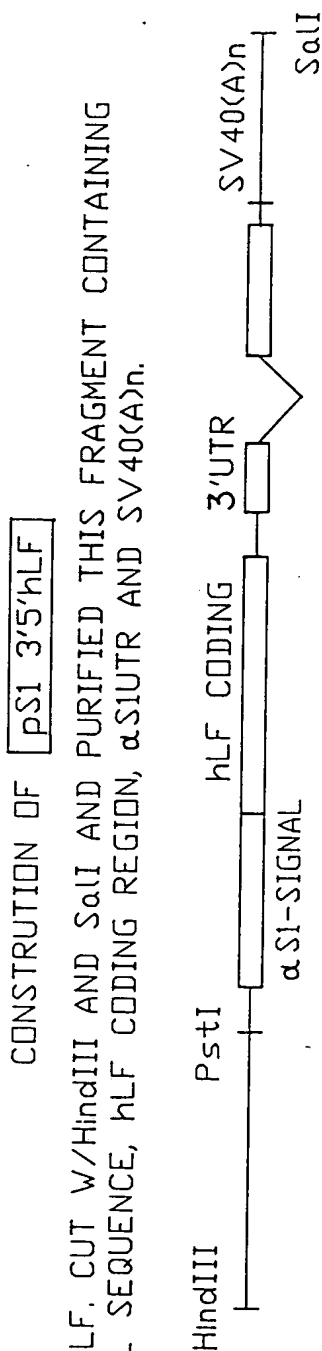
SUBCLONED INTO HindIII/SalI CUT pMHS1 5' FLANK TO YIELD:

PstI   αS1-SIGNAL   hLF CODING   KpnI   3'UTR   BamHI   SV40(A)n   SalI   NotI

HindIII   XhoI   HYBRID SPLICE   SIGNAL

NotI   NcoI   EXON1
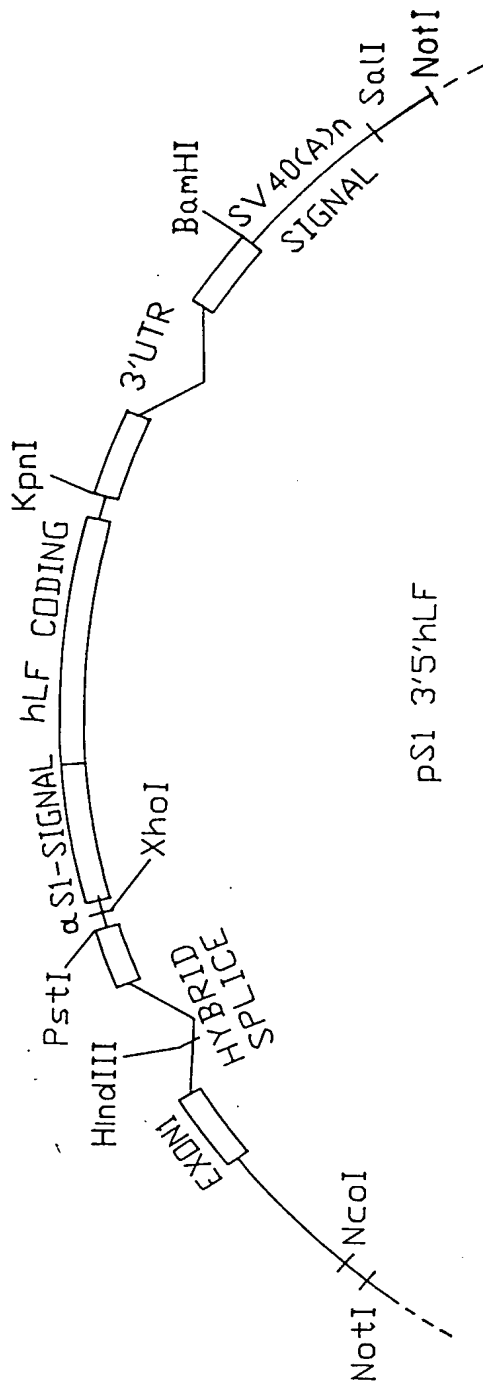
pS1 3'5'hLF

NOTE: pS1 5'hLF WAS MADE BY CUTTING pS13'5'hLF W/KpnI AND BamHI, FOLLOWED BY BLUNTING THE ENDS AND RELIGATING. THIS ELIMINATES THE SPLICED 3'UTR REGION.

FIG.-7E

COSMID cGP1hLF

p hLF3' 10kb

XhoI

hLF-CODING

αS1-3'FLANK

SalI

LINKER NotI

PHAGE GP1

BglII

XhoI

LINKER

16kb αS1 CASEIN

SalI
NotI

XhoI

XhoI

+

NotI

NotI

NotI

NotI

COSMID pWE15
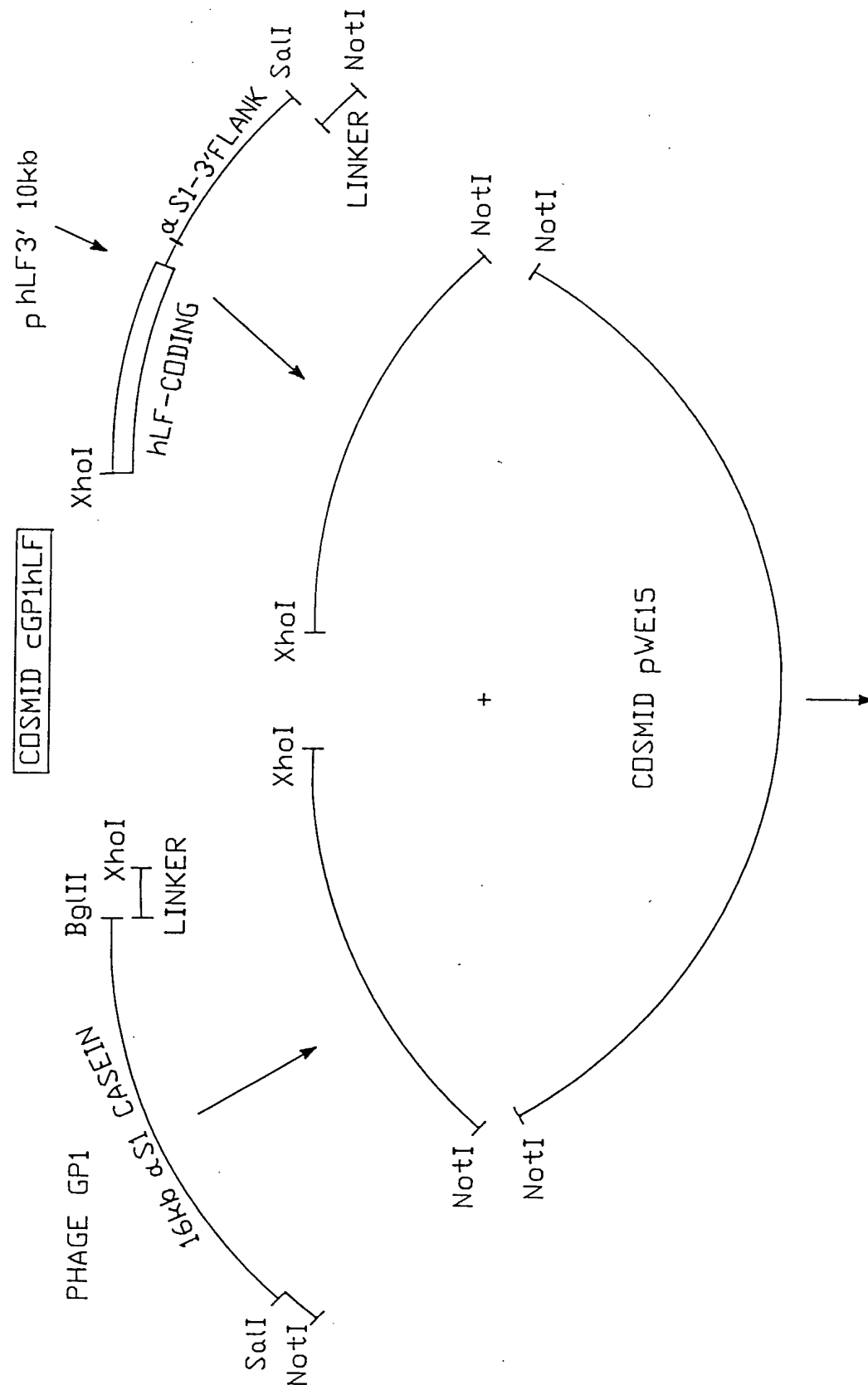
FIG.-7F

3 WAY LIGATION. THE DNA FROM THIS COSMID IS PREPARED BY CUTTING
WITH NotI AND PURIFYING THE EXPRESSION SEQUENCE PRIOR TO MICROINJECTION.
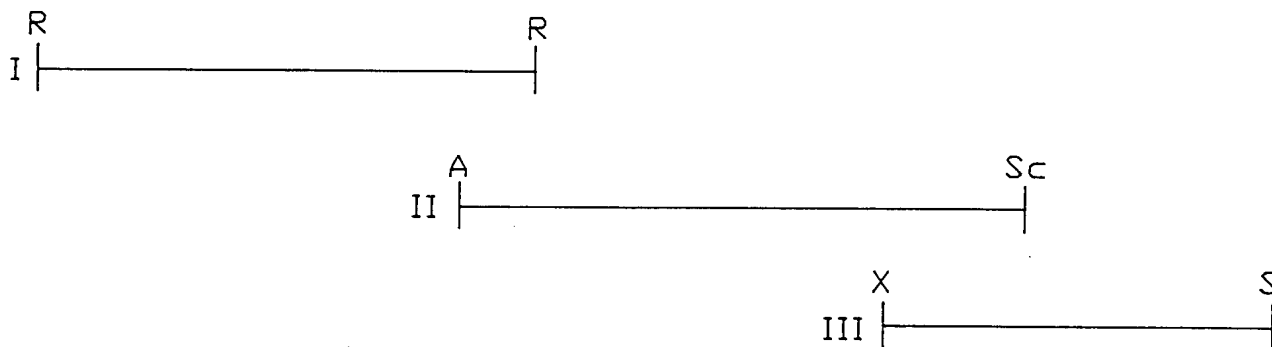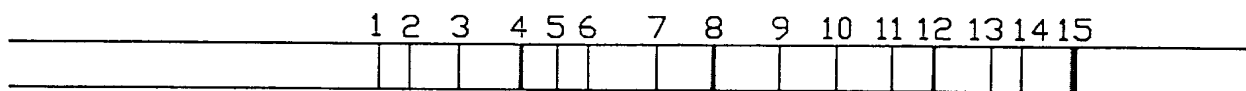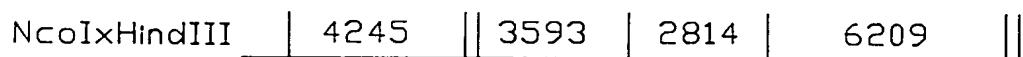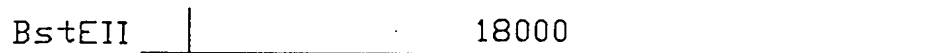
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

R   R
I |————————————————|

A   Sc
II |————————————————|

X   S
III |————————————————|

⊢—⊣
1kb

FIG.-8A

BstEII |————————————— 18000 —————————————|

NcoI | ——— 8079 ——— | ——— 9374 ——— ||

NcoIxHindIII | 4245 || 3593 | 2814 | 6209 ||

FIG.-8B
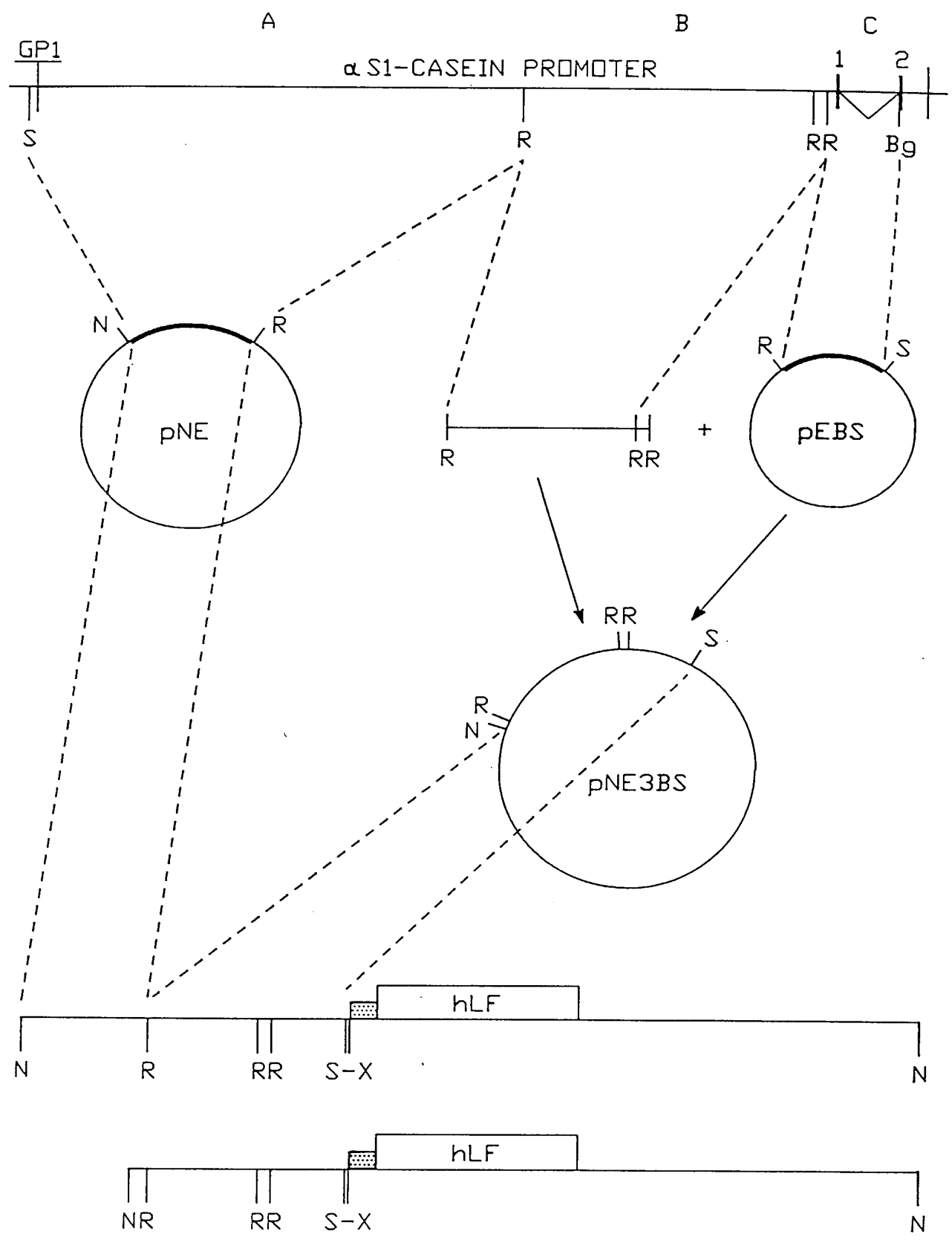
S=SalI R=EcoRI Bg=BglII N=NotI X=XhoI

FIG.-9

PHAGE CONTAINING PROTEIN C GENE

ClaI   NheI

FRAGMENT 2

INTRODUCED
ClaI SITE

(FRAGMENT AMPLIFIED FROM
PHAGE OR FROM HUMAN DNA)

ClaI   ClaI

+pGEM7A*ClaI

ClaI   ClaI
                              XbaI
                              SalI

pPCCC

CUT COMPLETELY WITH
XbaI AND ONLY PARTIAL
WITH ClaI AT SITE
INDICATED BY ARROW

                              ClaI

ClaI

                                            [NheI/XbaI]
                                            SalI

pPC
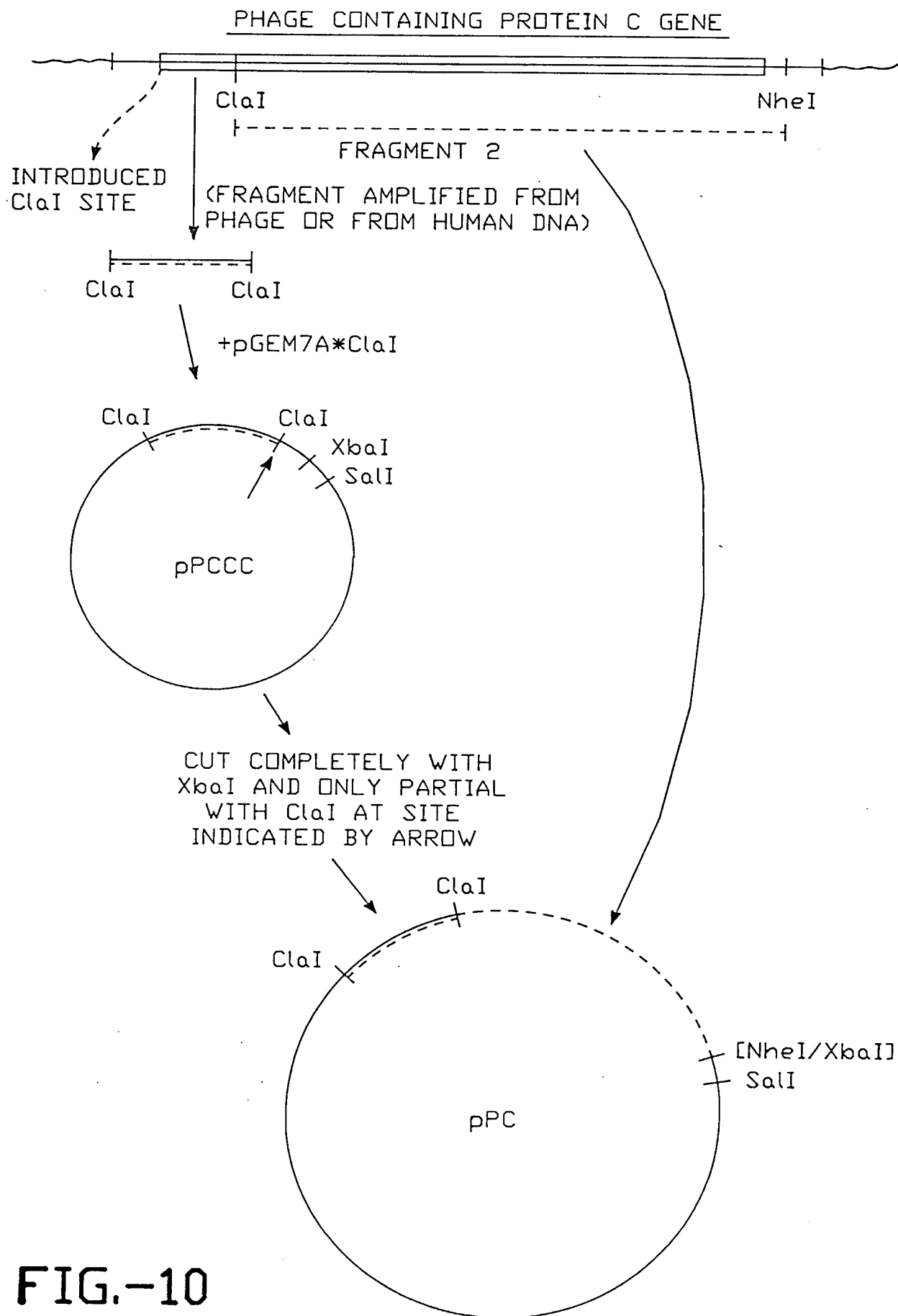
FIG.-10

5'- ATCACCTTGA TCATCAACCC AGCTTGCTGC TTCTTCCCAG

TCTTGGGTTC AAG gtattatgta tacatataac aaaatttcta tgattttcct ctgtctcatc

tttcattctt cactaatacg cagttgtaac ttttctatgt gattgcaagt attggtactt tcctatgata

tactgttagc aagcttgagg tgtggcaggc ttgagatctg gccatacact tgagtgacaa tgacatccac

tttgcctttc tctccacag GTGTCCACTC CCAGGTCCAA CTGCAG -3'

FIG.-11

START   STOP

N   R A   R R   R   A   R N

1 kb

EcoRI   5.8   0.3   3.8   8.0

Asp71M   13.3
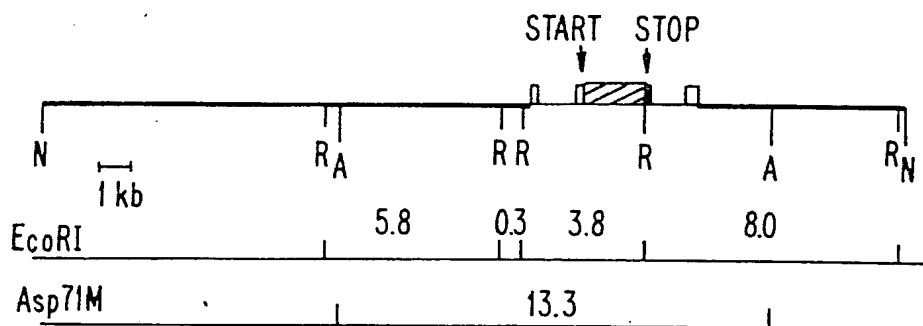
FIG. 12A.

COSMIDS CONTAINING ENTIRE hLF GENE (NOT DRAWN TO SCALE)

BamHI

13.2   ApaI ApaI   ClaI mluI BamHI   ClaI

5'   0.3 ApaI   3'

ClaI

13.1   ClaI

1.4 kb   2.1   2.7   2.7

7.5

COSMID VECTOR   NAME: C₂RB
                 SIZE: 8 kb

ClaI SITE DERIVED
FROM PLASMID
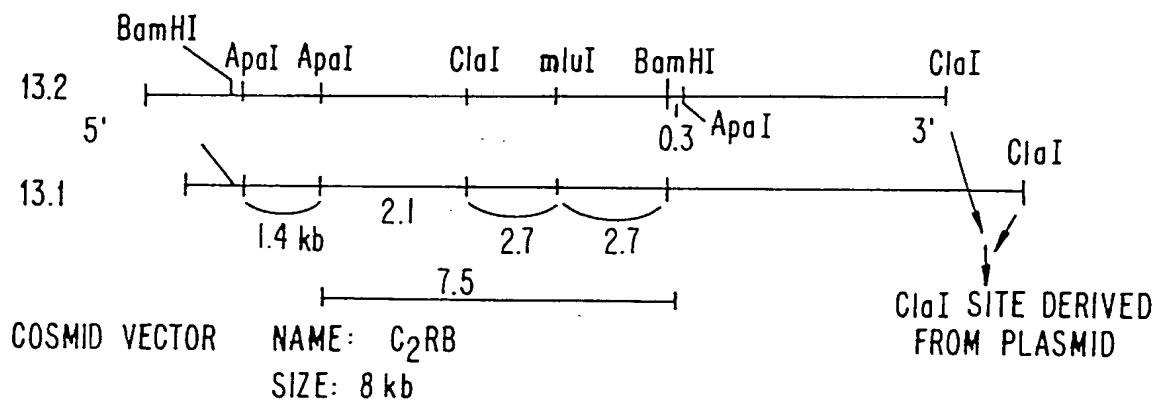
FIG. 13.

9 kb BamHI hLF FRAGMENT IN pUC19
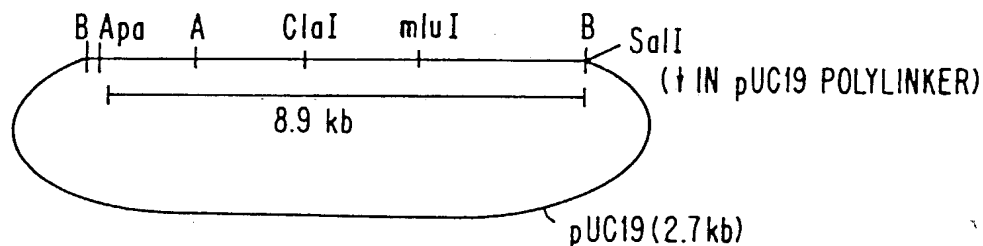
B Apa A   ClaI   mluI   B   SalI
                                 († IN pUC19 POLYLINKER)

8.9 kb

pUC19(2.7kb)

FIG. 14.
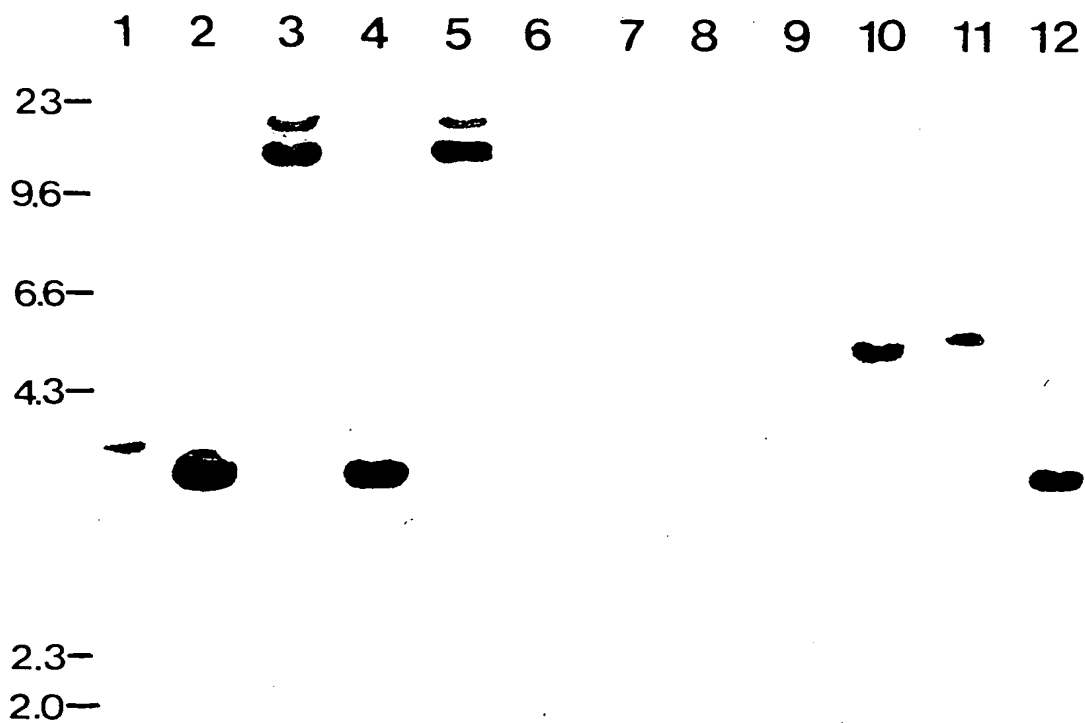
FIG. 12B.

LIGATION PRODUCT OF p16kbCS AND SYNTHETIC SEQ. (Cla-Apa)+ hLF FRAGMENT
(p8kbCS)

Not   αS₁ PROMOTER   Cla Apa Apa   Cla   mlu   Bam Sal I
          (16 OR 8 kb)

                                         8.9 kb Apa-Sal FROM hLF GENE

                                SYNTHETIC
                                SEQUENCE
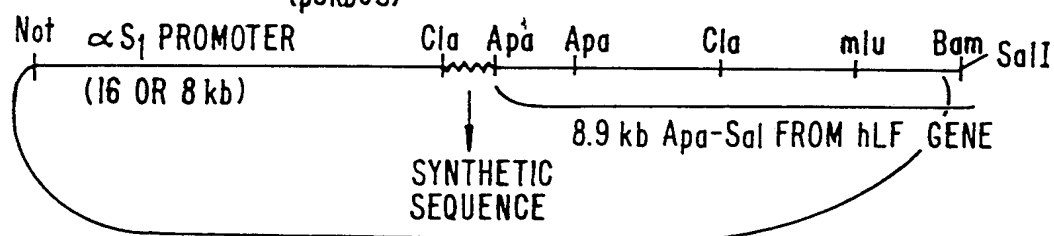
CLONING VECTOR: pkUN (4 kb)
CONSTRUCT NAME: 8 hLF gen 9k, OR 16 hLF gen 9k

FIG. 15A.

STRUCTURE OF ClaI-ApaI SYNTHETIC SEQUENCE

                                          TRANSLATION
                                          INITIATION CODON
ClaI
5'-CGA TAC CAA GTC GCC TCC AGA CCG CAG ACA TGA AAC TTG TCT
      T ATG GTT CAG CGG AGG TCT GGC GTC TGT ACT TTG AAC AGA

TCC TCG TCC TGC TGT TCC TCG GGG CC 3'
AGG AGC AGG ACG ACA AGG AGC C
                            NpaI
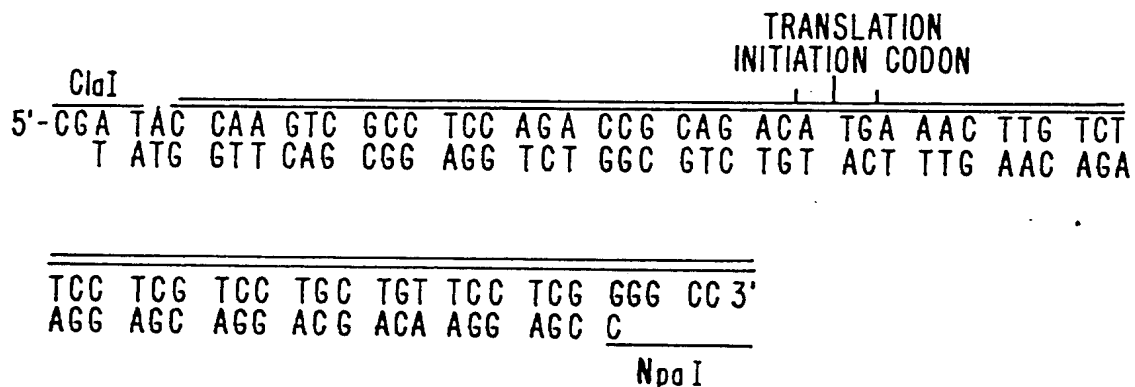
FIG. 15B.

+1: TRANSCRIPTION INITIATION SITE OF BOVINE αS₁ - CASEIN GENE

    +1
                 ClaI   hLF5'UTR   ApaI   hLF EXON 2
                                                INTRON 1

αS₁ CAS 5' UTR

                                    GGGCCCTCG   GACTGTGTCTGGCT
                                    ApaI   14 bp
                                              INTRON 1
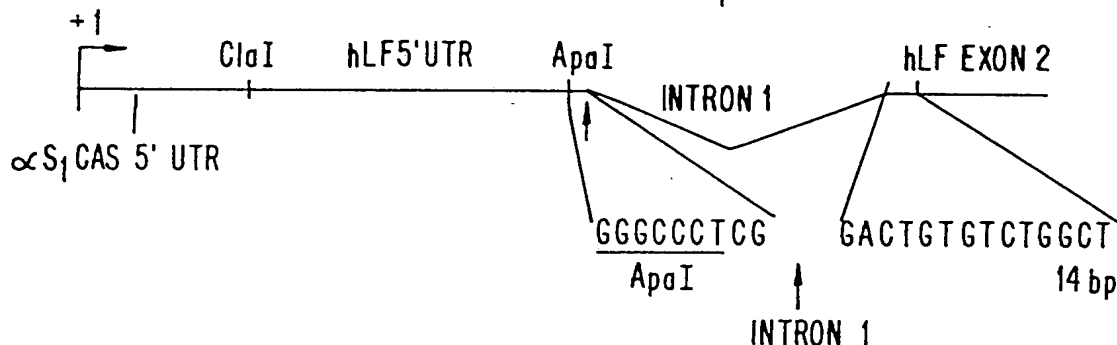
STRUCTURE OF REGION CONTAINING EXON 1 (HYBRID αS₁-CASEIN/hLF EXON) AND
PART OF EXON 2 OF THE GENOMIC hLF CONSTRUCTS DEPICTED IN FIGS. 15A THROUGH
17.

FIG. 15C.

COINJECTION

NotI   Cla A A   ClaI mlu   SalI   (SEE FIG. 15A)   ClaI SITE: DERIVED FROM
                                                            THE CLONING VECTOR

Cla I   +

OVERLAP: 5.4 kb

ClaI   (SEE FIG. 13)
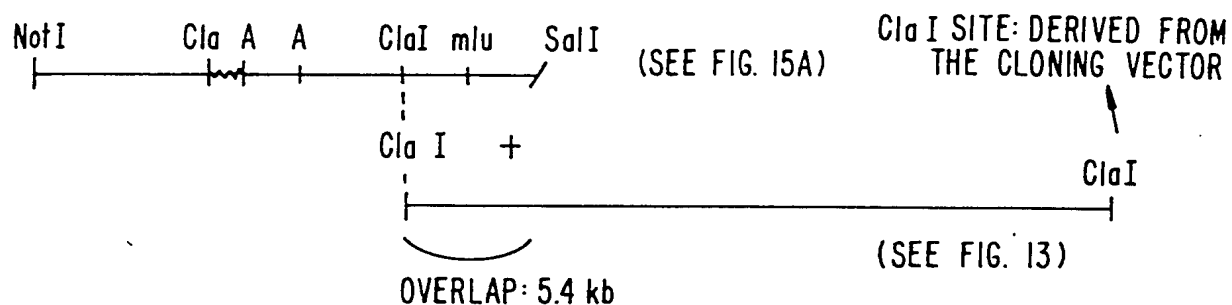
FIG. 16.

GENERATION OF 8 hLF GENE

NotI   Cla A   A Cla mluI   ← FROM 8hLF gen 9k
                      1.4 2.1 2.7

+   mluI   26-27 kb   Cla I
                (FROM 13.2)

+   Cla/Not

↓
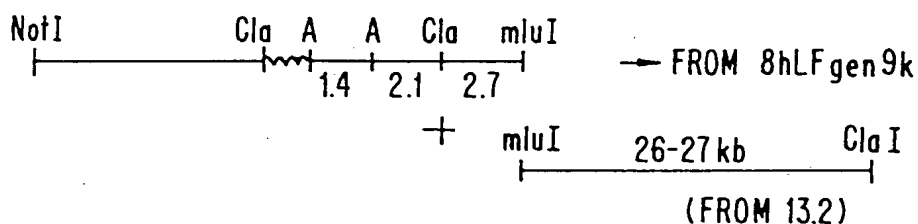
LIGATE INTO Not I CUT COSMID

Cla/Not LINKER: SalI

C G G T C G A C A T C G A T G C
  C A G C T G T A G C T A C G C C G G
                    ClaI   NotI

ClaI

(DESTROYS ClaI SITE)

FIG. 17.

λ BLG-13
BamHI = B
Hind III = H
E = EcoRI
S = SacI

B H   H   B   E   B   BE H   B   B

EXON I

EXON VII (LAST EXON)

= PUBLISHED MAP

Not I — E   S   S   E 4.3 kb
SmaI   SfiI

pkUN S

Not I — SacI DIGEST = 3.2 kb FRAGMENT

NotI   SacI   SacI

+   LINKER:

Pvu II   Cla I

CAG AAG CGT GAC CCC AG T ATC   GAT ACC TGG
TCGAGTC TTC GCA CTG GGG TC A TAG   CTA TGG ACC AGCT
SacI   SalI
OVERHANG   OVERHANG

BLG SEQUENCE   "RANDOM" SEQUENCE

LIGATE INTO pkUN△C →

Not   S   S Pvu II

Cla Sal

pBLG3.2

FIG. 18.

30

Not   Sac   Sac PvuII
                               Cla
                               Sal

pBLG3.2, Cla + Sal CUT

+

Cla Apa Apa Cla   mlu   Bam Sal

8.9 kb hLF FRAGMENT, Cla - Sal,
FROM 8 hLF gen 9k

LIGATE

Not   Sac   Sac PvuII
                    Cla Apa Apa Cla mlu   Bam Sal   pBLGhLFgen9k

        BLG   hLF

                                        —— ISOLATE 9.9 kb Not-mlu FRAGMENT
Not   9.9 kb   mlu
                                  +

                    mlu   23-24 kb   Not
                            (DERIVED FROM 8hLF gen)

                LIGATE

Not   mlu   Not

    BLG   -   hLF

FIG. 19.

FIG. 20.

DESIGN OF 16, 8 hLZ

: EXON FROM αS_1 - CASEIN GENE (NON-CODING)

: EXON FROM hLZ GENE

+1 : TRANSCRIPTION INITIATION SITE

: NON-CODING PART OF hLZ EXON 4

1, 8 AND 9 : αS_1 - GENE EXON NUMBERS

I - IV : NUMBERS OF hLZ EXONS

FLANKING REGIONS: FROM αS_1 CASEIN GENE

  (SIGNAL SEQUENCE: hLZ)

FIG. 21.

DESIGN OF 16, 8 hLZ3

LEGEND: SEE FIGURE 9

HYBRID INTRON AND SPLICE ACCEPTOR SITE

(SIGNAL SEQUENCE: hLZ)

FIG. 22.

λ 7.2.1

λ-ARM   I   II   III   IV   λ-ARM

Hinc II   Sal I

I - IV : hLZ EXON 5

▨ : EXON 4, NON-CODING

5.3 kb

Hinc II   Sal I

PRIMERS:

+3: 5'UTR hLZ

CATCGATCCCTAGCACTCTGACCTAGCAGTC

CATG GTAGCTAGGGATCGTGAGACTGGATCGTCAG

KpnI   Cla I   HincII

LIGATE 5.3 HincII-Sal I } →
Kpn - HincII
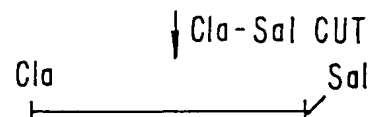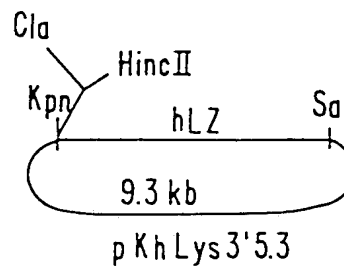
Cla
HincII
Kpn
hLZ   Sal

9.3 kb

pKhLys3'5.3

↓ Cla-Sal CUT

Cla   Sal

FIG. 23A.

Cla   hLZ   Sal

+

Not   Cla   Sal
αS₁ PROMOTER
(0.7kb)

LIGATE

p - 0.7 kb CS

Not   αS₁   Cla   hLZ   Sal
PROMOTER

CUT   Sal I
+

αS₁ CAS 3'UTR 3'
5'

XhoI   Sal I
NotI   NotI

pKE 3' E₁₀   ISOLATE   XhoI - Sal I

LIGATE

NotI   ClaI   hLZ   Sal I*/XhoI*   Sal I
αS₁ PROMOTER   5.3kb   3'UTR αS₁
0.7 kb   8 kb

p0.7,8 hLZ
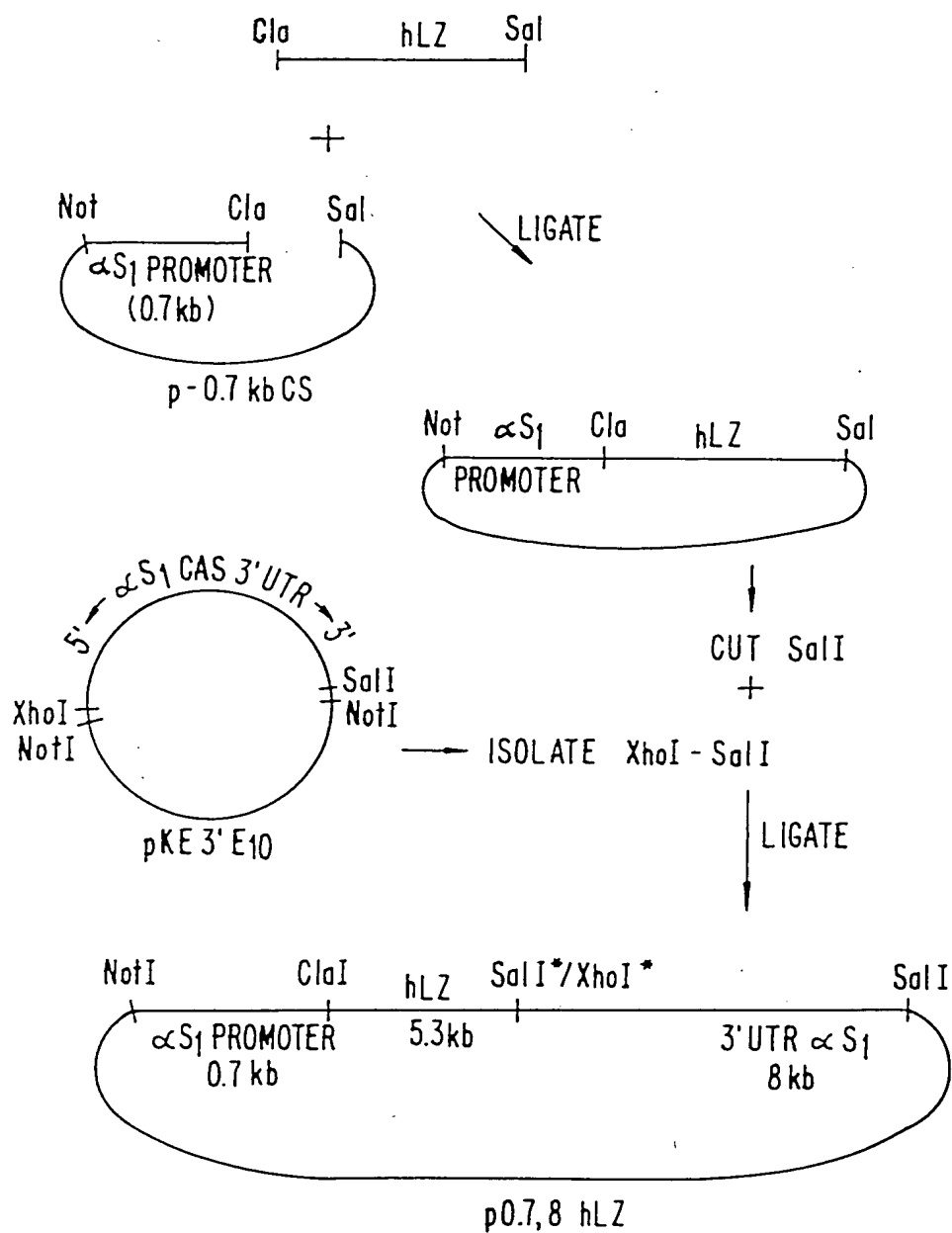
FIG. 23B.

LINKER S_1/S_2:

SalI$^\circledast$ – NotI – SalI$^\oplus$

$\circledast$: ▪ DESTROYED SITE

5'
T C G A G $\overline{C\,G\,G\,C\,C\,G\,C}$ C G G A C C G G G C C G C C T C G →
        C G C C G G C G G C C T G G C C C G G C G G A G C →

SalI$^\circledast$ | NotI

                 3'
→ G C C T C G C G A
→ C G G A G C G C T | A G C T
                  └ SalI$^\circledast$

FIG. 23C.

CUT p0.78 hLZ SalI
↓

Not   Cla   Sal*/Xho*   SalI

$\propto S_1$ PROMOTER   hLZ

↓ LIGATE $S_1 S_2$

Not   Cla   Sal*/Xho*   NotI

p0.78 hLZNt   Cla-Not CUT
                                   ↓

ClaI   Not I

13.3 kb

FIG. 23D.

Not   αS₁-PROMOTER   Cla   Sal

p-16 CS

CUT Cla + Sal

NotI   ClaI

14.5 kb

+

ClaI   hLZ   3' αS₁   NotI

LIGATE IN pWE15

NotI   5' αS₁   Cla   hLZ   3' αS₁   NotI

p16, 8 hLZ
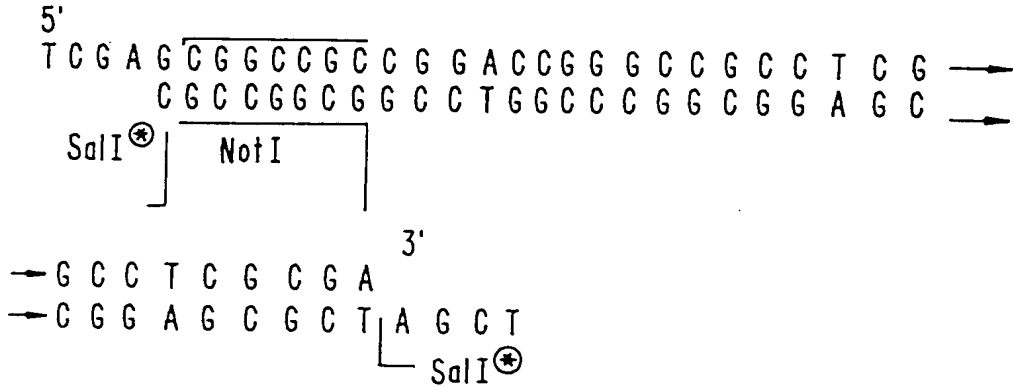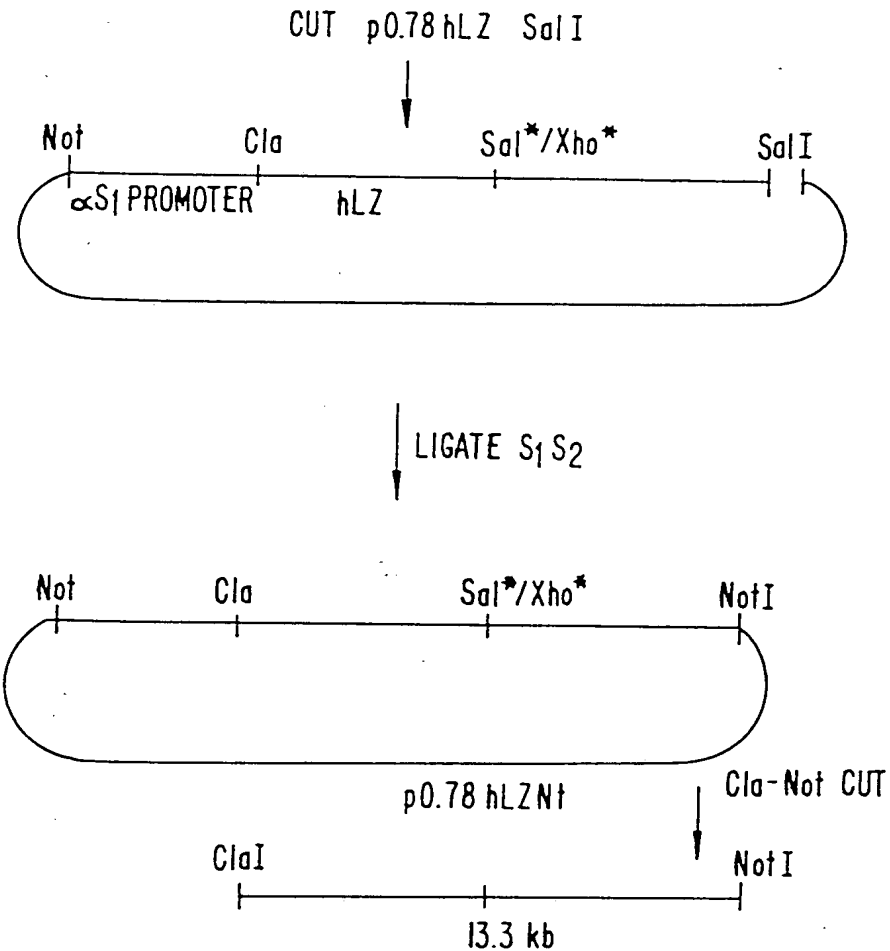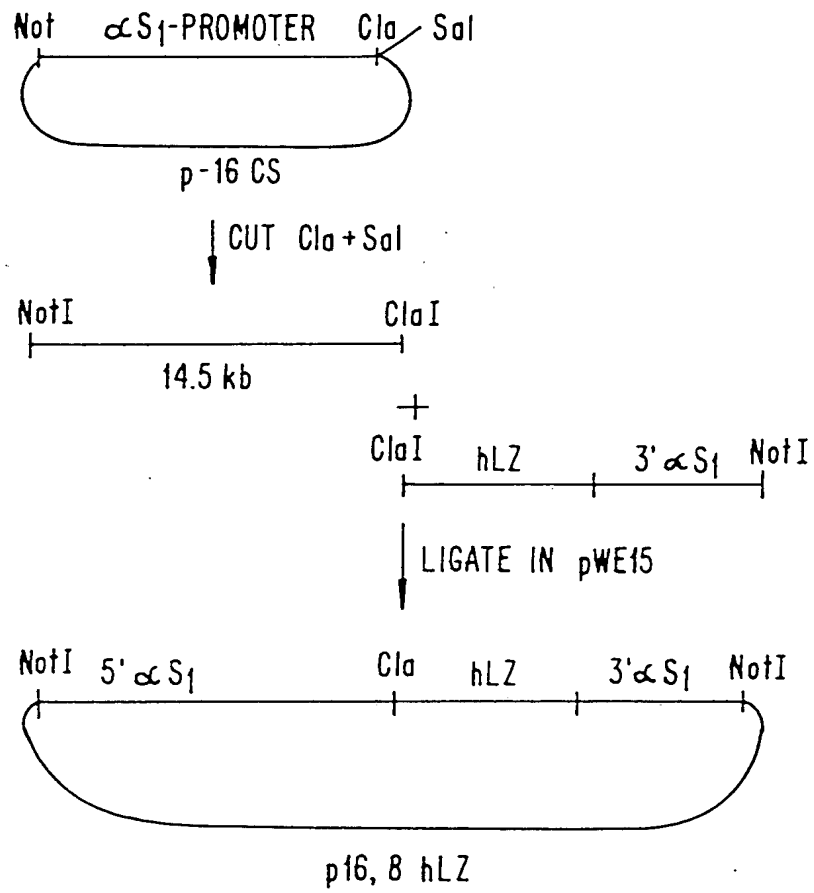
FIG. 23E.

```
BOVINE 2 GGAAGTGCCTGGAGA...TTAAAATGTGAGAGTGGAGT....GGAGGTTG 44
          ||||||||| |  | || |||||||||||||||||        |||||||
SHEEP 84 GGAAGTGTCCTGGGAGATTTAAAATGTGAGAGGCGGGAGGTGGGAGGTTG 133

     45 GGTCCTGTAGGCCTTCCCATCCCACGTGCCTCACGGAGCCCTAGTGCTAC 94
         ||  ||||||  ||||| |||||||||||||||| |  |||  |||||| |
    134 GGCCCTGTGGGCCTGCCCATCCCACGTGCCTGCATTAGCCCCAGTGCTGC 183

     95 TCAGTCATGCCCCGCAGCAGGGGTCAGGTCACTTTCCCATCCTGGGGGT 144
         |||| | ||||||||| ||||||||||||||||||||||| |||| |||||
    184 TCAGCCGTGCCCCCGCCGCAGGGGTCAGGTCACTTTCCCGTCCT.GGGGT 232

    145 TATTATGACTGTTGTCATTGTTGTTGCCATTTTTGCTACCCTAACTGGGC 194
         |||||||||| | |||||||| ||||||| |||||||||||||||||||||
    233 TATTATGACTCTTGTCATTGCCATTGCCATTTTTGCTACCCTAACTGGGC 282

    195 AGCGGGTGCTTGCAGAGCCCTCGATACTGACCAGGTTCCCCCCTCGGAGC 244
         |||  |||||||||||||||||||||||| |||||||   ||||||||||
    283 AGCAGGTGCTTGCAGAGCCCTCGATACCGACCAGG.TCCTCCCTCGGAGC 331

    245 TCGACCTGAACCCCATGTCACCCTCGCCCCAGCCTGCAGAGGGTGGGTGA 294
         ||||||||||||||||||||||||||| |||||||||||||||||||||| |
    332 TCGACCTGAACCCCATGTCACCCTTGCCCCAGCCTGCAGAGGGTGGGTGA 381

    295 CTGCAGAGATCCCTTTACCCAAGGCCACAGTCACATGGTTTGGAGGAGAT 344
         |||||||||||||||||| |||||||||| |||||||||||||||||||| |
    382 CTGCAGAGATCCCTTCACCCAAGGCCACGGTCACATGGTTTGGAGGAGCT 431

    345 GGTGCCCAAGGCAGAAGCCACCCTCCA.GACACACCTGCCCCCAGTGCTG 393
         |||||||||||||||  |||||||||||  |||||||||  ||||||||||
    432 GGTGCCCAAGGCAGAGGCCACCCTCCAGGACACACCTGTCCCCAGTGCTG 481

    394 GCTCTGACCTGTCCTTGTCTAAGAGGCTGACCCCAGAAGTGTTCCTGGCG 443
         ||||||||||||||||||||||||||||||||||| |||||||||||||||
    482 GCTCTGACCTGTCCTTGTCTAAGAGGCTGACCCCGGAAGTGTTCCTGGCA 531

    444 CTGGCAGCCAGCCTGGACCCAGAGCCTGGACACCC.CCTGCGCCCCCACT 492
         |||||||||||||||||||||||||| |  |||||||| |||| ||||| ||
    532 CTGGCAGCCAGCCTGGACCCAGAGTCCAGACACCCACCTGTGCCCCCGCT 581

    493 TCTGGGGGCGTACCAGGAACCGTCCAGGCCCAGA..GGGCCTTCCTGCTT 540
         ||||||  || |||||||||||||| |||||||| |  ||| |||||||||
    582 TCTGGGGTC.TACCAGGAACCGTCTAGGCCCAGAGGGGGACTTCCTGCTT 630

    541 GGCCTCGAATGGAAGAAGGCCTCCTATTGTCCTTCGTAGAGGAAGCAACC 590
         ||||| | ||||||||||||||||||||||||| |||||||||||| |||
    631 GGCCTTGGATGGAAGAAGGCCTCCTATTGTCC.TCGTAGAGGAAGCCACC 679

    591 CCAGGGCCCAAGGATAGGCCAGGGGGGATTCGGGGAACCGCGTGGCT.CC 639
         || |||||| ||||| |||| ||||||||| ||| ||||||||||||  ||
    680 CCGGGGCCTGAGGATGAGCCAAGTGGGATTCCGGGAACCGCGTGGCTGGG 729

    640 GGCGCGGCCCGGGCTGGCTGGCTGGC..CCTCCTCCTGTATAAGGCCCCG 687
         ||| | |||||||||||||||||||  || |  | |||||||||||||||||
    730 GGCCCAGCCCGGGCTGGCTGGCCTGCATGCGCCTCCTGTATAAGGCCCCA 779
```

FIG. 24A.

688 AGCCCG.CTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCGTGACC 736

780 AGCCTGCCTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCACGACC 829

737 CCAGCTGCAGCC(ATG)AAGTGCCTCCTGCTTGC......CCTGGCCCTCAC 780

830 CCAGCTGCAGCCATGAAGTGCCTCCTGCTTGCCCTGGGCCTGGCCCTCGC 879

781 CTGTGGCGCCCAGGCCCTCATCGTCACC 808

880 CTGTGGCGTCCAGGCCATCATCGTCACC 907

*TRANSLATION INITIATION CODON

## FIG. 24B.

LINKER GP 278/279

TRANSLATION START SITE (αS₁ SIGNAL SEQUENCE)

ClaI

CGATAACCATGAAACTTCTTATCCTCACCTGTCTTGTGGCTGTTGCTCTTG

TATT — ETC.

αS₁ ─┬─ hLZ SEQUENCE

BalI

—CCAAGGTCTTTGAAAGGTGTGAGTTGC

ETC.—AACC

FIG. 25.

CONSTRUCTION OF <u>16.8 A hLZ3</u>:

NotI   αS₁   ClaI

SalI

p0.78 hLZ
x ClaI/SalI
(4.7 kb)

→ LIGATE TO GP 278/279

(= ClaI – BalI)

ClaI ⌇⌇⌇ BalI

ClaI   BalI

SalI

BalI   5.3 kb   SalI

HincII BalI BalI BalI   hLZ   SalI
         0.7 0.9   3.7 kb

pKhLys 3'5.3

LIGATE

NotI   ClaI BalI   SalI

ClaI   p0.7AhLZΔ3'

⌇⌇⌇   SalI

LIGATE

NotI   ClaI SalI

14.5 kb
αS₁ PROMOTER

NotI   8 kb 3' FLANK
                    αS₁

DERIVED FROM
16.8 hLF3

NotI 5'αS₁ ClaI   hLZ   SalI 3'αS₁ NotI

p16.8AhLZ3
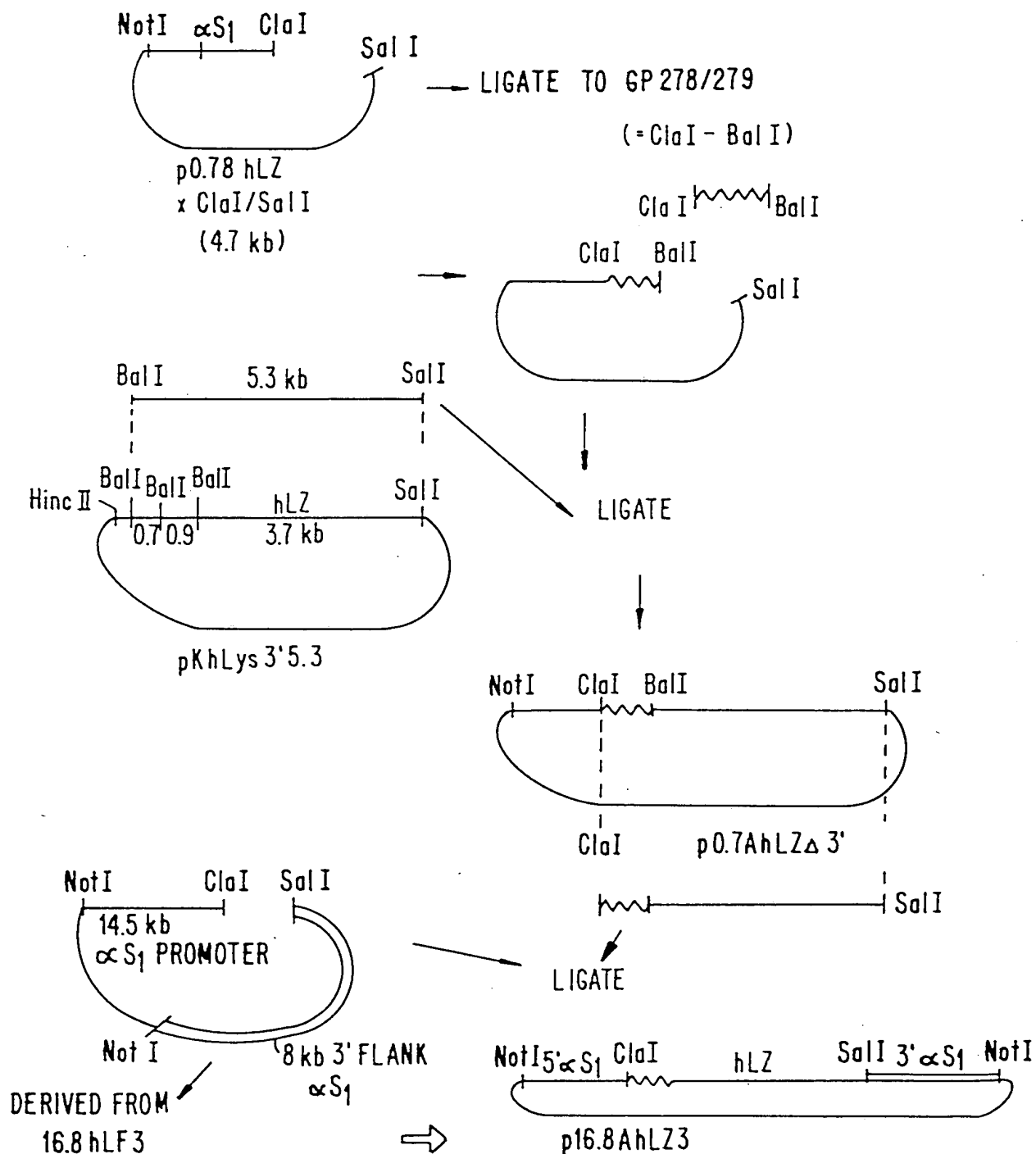
FIG. 26.

CONSTRUCTION OF 16 A hLZ3:

FIG. 27.